

Accounting for Relatedness in Family Based Genetic Association Studies

P.F. McArdle^a J.R. O'Connell^b T.I. Pollin^b M. Baumgarten^a A.R. Shuldiner^{b,c}
P.A. Peyser^d B.D. Mitchell^b

^aDepartment of Epidemiology and Preventive Medicine, ^bDivision of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, ^cGeriatric Research and Educational Clinical Center (GRECC), Veterans Administration Medical Center, Baltimore, Md., and ^dDepartment of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Mich., USA

Key Words

Genetic association studies · Type 1 error · Variance components · Regression modeling

Abstract

Objective: Assess the differences in point estimates, power and type 1 error rates when accounting for and ignoring family structure in genetic tests of association. **Methods:** We compare by simulation the performance of analytic models using variance components to account for family structure and regression models that ignore relatedness for a range of possible family based study designs (i.e., sib pairs vs. large sibships vs. nuclear families vs. extended families). **Results:** Our analyses indicate that effect size estimates and power are not significantly affected by ignoring family structure. Type 1 error rates increase when family structure is ignored, as density of family structures increases, and as trait heritability increases. For discrete traits with moderate levels of heritability and across many common sampling designs, type 1 error rates rise from a nominal 0.05 to 0.11. **Conclusion:** Ignoring family structure may be useful in screening although it comes at a cost of an increased type 1 error rate, the magnitude of which depends on trait heritability and pedigree configuration.

Copyright © 2007 S. Karger AG, Basel

Introduction

Susceptibility genes for complex diseases have been difficult to identify and it has been suggested that successful discovery of such genes will require well-designed epidemiological studies [1, 2]. To that end, genetic association studies have become an important analytic strategy to help identify the genetic determinants of complex diseases, and in this context, study designs that include both probands and family members are common. Moreover, the growing affordability of high throughput genotyping has increased the scope of such studies, and genome wide association studies involving hundreds of thousands of single nucleotide polymorphisms (SNPs) are now underway [3–5]. The analysis of such high volume genotype data requires the development of appropriate analytic strategies that are both scientifically appropriate and operationally manageable.

Successful mapping of genes requires suitable modeling of genetic effects. Such modeling is often performed within a regression framework where the effect of genotype on phenotype can be assessed with consideration of covariate effects and interactions. Much recent work has focused on developing tractable methods for such modeling when very large numbers of SNPs are to be analyzed or genotyping is done in two phases [6, 7]. However, the

modeling strategy becomes more complicated when the study sample includes families and the trait is correlated among related individuals by virtue of their shared genetic background.

Regression is a robust and well understood modeling technique that allows for flexibility in testing a variety of hypotheses [8]. In genetic epidemiology, regression modeling is commonly used to test for association among unrelated individuals and for association in the presence of linkage with family data [9, 10]. Simple regression models assume that each observation is independent. When this is not the case, as for example when multiple members of a family are studied, statistical techniques should be used to account for the nonindependence. Variance components modeling is one approach for accounting for correlations among related individuals [e.g. 11]. Other commonly used approaches include modeling of family unit as a random effect using a mixed model and general estimating equations [12–14].

While accounting for familial relationships in the analysis is desirable in terms of providing appropriate estimates of the model parameters, many of the current versions of available software products designed to analyze large genome wide association datasets use statistical methodologies that ignore correlation of phenotypes among related individuals. Additionally, the number of total models to be analyzed can be extraordinarily large in genome wide analyses, particularly when there are multiple outcomes of interest for each SNP and multiple models to be considered (e.g., multiple genetic models and different combinations of covariates and interaction terms). Using simplified analytic models that ignore family structure might be an appealing alternative if computational burden is reduced and if model performance can be identified in advance. In this regard it would be helpful to know the relative impact of ignoring familial relatedness on such features as study power, type 1 error, and parameter estimates.

The purpose of this paper is to provide empirical assessments of the effect of ignoring familial relatedness using regression based association tests of genetic data across a range of settings. Specifically, we consider an analysis to evaluate the association between a SNP (or other bi-allelic variant in the genome, including insertions or deletions) and a disease outcome when the study subjects are related. The effect of familial relatedness can be expected to be based primarily on how related the individuals in the study are and how correlated the trait is among related family members. To assess the impact of these factors, we simulated discrete phenotypes across a

range of heritabilities that were associated with SNPs having a variety of effect sizes in four different types of pedigree configurations: sib pairs, large sibships, nuclear families, and extended families. Additionally, we simulated data using ‘real world’ pedigrees derived from an ongoing family study of diabetes [15]. For each family structure and across a range of heritabilities, we analyzed the simulated data both ignoring and accounting for familial relatedness. We then compared parameter estimates, power, and type 1 error rates between the two modeling approaches.

Methods

The overall goal of this study was to assess the impact of accounting for family structure on the performance of strategies for detecting genetic associations in family based studies. This was accomplished by analysis of simulated datasets. First genotype and phenotype data were simulated in a variety of family structures. Then each simulated dataset was analyzed twice, once accounting for family structure and once ignoring family structure, and the results of the two analytic methods were compared.

Four different pedigree configurations were considered, ranging from small families with few relative pairs to larger families with more relative pairs. The four pedigree configurations were sib pairs, large sibships, nuclear families, and extended families. Each configuration of pedigrees was constructed to have 500 individuals for whom phenotype-genotype associations were assessed. Figure 1 describes the structure of each pedigree configuration. The sib pair configuration included 250 independent sibships of size 2. The large sibship configuration included 125 independent sibships of size 4. The nuclear family configuration included 125 families consisting of 2 offspring and both parents. The extended family configuration included 25 families of 20 members each, comprising a sib trio, their spouses, and 4 children from each of the sib-spouse pairs. Thus, this configuration leads to 100 parent pairs, 525 sib pairs, 750 parent-offspring pairs, 600 avuncular pairs, and 1200 first cousin relationships.

In addition to the pedigree structures described above, we performed simulations in a set of pedigrees of participants from one of our ongoing studies of the genetics of type 2 diabetes, the Amish Family Diabetes Study (AFDS) [15]. This sample included 511 phenotyped individuals, 150 of whom were classified as cases (subjects with diabetes in the parent study) and 361 as controls (subjects with normoglycemia in the parent study). For purposes of the current study, these individuals were merged, where possible, into two-to-three generation families, with the resulting pedigree configuration consisting of 126 singletons and 64 pedigrees connecting two or more study subjects (i.e., 126 individuals in these truncated Amish ‘analysis’ pedigrees were unrelated to any other study subject). The 64 pedigrees ranged in size from 2 to 22 phenotyped individuals. Figure 1 shows the pedigree structure for three sample families in the AFDS. For the study described in this report, data were simulated only for subjects phenotyped in the parent study so as to mimic the original missing data patterns. The individuals for whom data were simulated in-

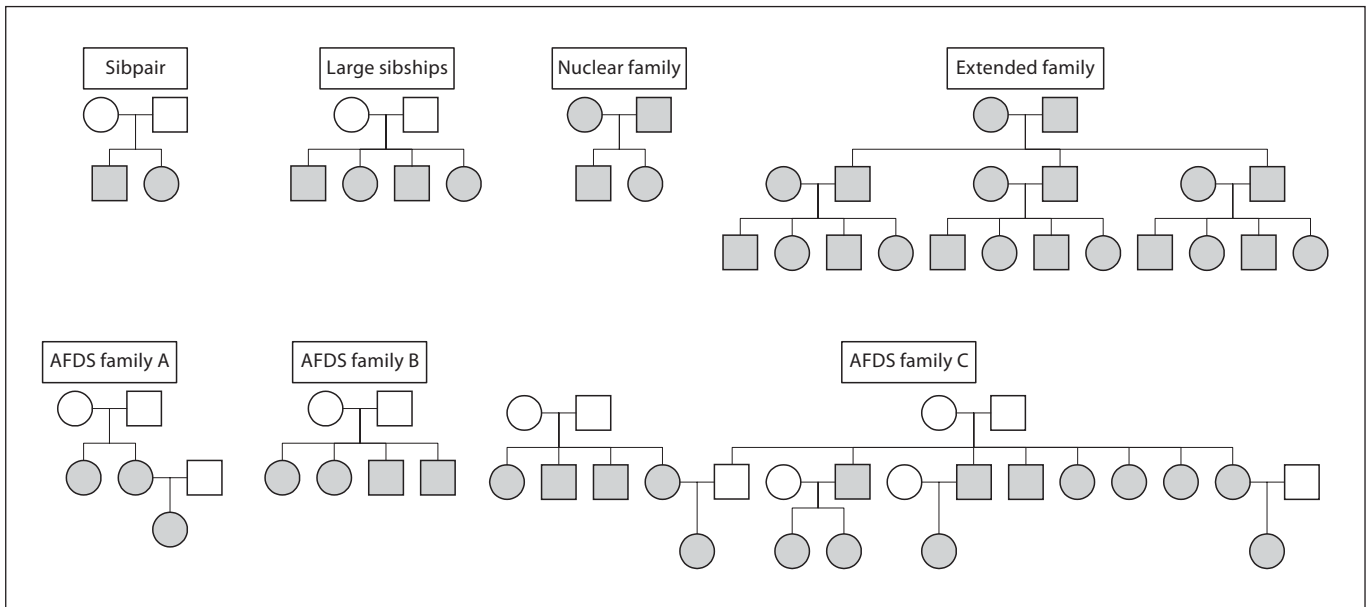


Fig. 1. Example of pedigrees of families used for simulations. Genotype and phenotype data were analyzed for shaded individuals only.

cluded 528 sib pairs, 180 parent-offspring pairs, 347 avuncular pairs, and 156 first cousin pairs.

For purposes of this study, quantitatively distributed traits were simulated based on assigned genotypes and the traits were then converted to a discrete trait (e.g., affection status) based on threshold values. For purposes of presentation and ease of interpretation, the QTL effects are described in terms of the odds ratio. Where appropriate, comments are provided in the text regarding the comparability of conclusions obtained from analysis of the converted discrete traits versus those obtained through analysis of the original quantitatively distributed traits.

Simulation Approach

We simulated a biallelic quantitative trait locus (QTL) assuming Hardy-Weinberg equilibrium under an additive model with a polygenic component using the SOLAR simqtl option (Southwest Foundation for Biomedical Research, San Antonio, Tex., USA). The minor allele frequency ranged from 0.1 to 0.5 in increments of 0.1 and the residual heritability of the polygenic component ranged from 0.1 to 0.9 in increments of 0.1. We converted the simulated quantitative trait to a discrete trait using a liability threshold assuming a 30% population prevalence of being affected. Individuals with a simulated trait value equal to or above the threshold were considered affected and those with values below the threshold as unaffected. The odds of being affected was calculated for each genotype using the cumulative normal distribution as the percent area of the distribution above the threshold. The odds of being affected for each of the genotypes were used to parameterize the effect size as an odds ratio. Six effect sizes were simulated (odds ratio = 1.0, 1.2, 1.4, 1.6, 1.8, 2.0). An additive genetic model was simulated so that the effect is linear on the odds ratio scale.

In the following text and tables, we have presented the heritability of the simulated underlying quantitative trait, not the heritability of the analyzed discrete trait. In some cases, the realized heritability of the discrete trait differed slightly from that of the underlying quantitative trait. Specifically, the realized heritability of the discrete trait tended to be slightly greater than that of the quantitative trait for the sib pair design, identical to that of the quantitative trait for the large sibship and nuclear family designs, and slightly less than that of the quantitative trait for the extended family design.

Analytic Approach

Analysis was performed using a variance components discrete trait model under an additive genetic model. Statistical significance of the genetic effect was determined by likelihood ratio test with one degree of freedom, in which we compared the likelihood of a full model, where the genetic effect was estimated, to that of a restricted model in which the genetic effect was constrained to be zero. The relatedness among family members was accounted for by including the pairwise kinship matrix specifying the degree of relatedness between each pair of individuals in the analysis set. The variance components approach models the correlations between the simulated trait (the dependent variable) and genotype (independent variable), conditional on the residual correlations among individuals implied by the pedigree structure. Specifically, the covariance between each pair of individuals within the pedigree is estimated as a function of their degree of relationship, the trait heritability, and the phenotypic variance of the trait. The model is thus defined as:

$$Y = X\beta + g + e,$$

Table 1. Recovery of the simulated SNP effect when accounting for and ignoring family structure for a trait with background heritability of 0.40

Simulated odds ratio	Simulated odds ratio			
	sib pair pedigrees		extended family pedigrees	
	account for family	ignore family	account for family	ignore family
1.0	1.00 (0.76–1.34)	1.00 (0.75–1.34)	1.00 (0.73–1.38)	1.00 (0.72–1.40)
1.2	1.20 (0.91–1.60)	1.21 (0.91–1.61)	1.20 (0.87–1.65)	1.20 (0.85–1.69)
1.4	1.40 (1.05–1.90)	1.41 (1.06–1.91)	1.40 (1.01–1.98)	1.40 (1.00–1.98)
1.6	1.59 (1.20–2.18)	1.61 (1.20–2.19)	1.60 (1.16–2.24)	1.60 (1.13–2.27)
1.8	1.79 (1.33–2.46)	1.80 (1.33–2.48)	1.81 (1.28–2.59)	1.81 (1.27–2.60)
2.0	1.98 (1.46–2.72)	1.99 (1.47–2.76)	2.00 (1.44–2.86)	2.00 (1.43–2.90)

Mean recovered odds ratios are given with the empirical 95% confidence intervals for 5,000 simulations.

Each effect size was simulated in every pedigree structure 5000 times. Empirical confidence intervals were defined by the point where 125 simulations (2.5%) returned point estimates less than the lower confidence limit given. Similarly, 125 simulations (2.5%) returned point estimates greater than the upper confidence limit.

where Y is a vector of individual phenotype values, and X is a design matrix accommodating an intercept and a vector of individual genotype values coded as 0, 1 or 2 representing the number of copies of the at-risk allele. β is a vector containing the intercept and the SNP effect of interest. The g term is the polygenic component that is distributed multivariate normally with a mean of zero and a covariance equal to two times the kinship matrix times the expected variance due to the additive effect of genes. The e term is a normally distributed error component with mean zero. Generalized least squares estimates of the parameters of interest are given by:

$$\beta = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

$$\text{var}(\beta) = (X^T V^{-1} X)^{-1}$$

where V is the variance-covariance matrix and is a function of residual trait heritability and the relationships implied by the pedigree structure. The likelihood of the observed data was computed under the assumption of multivariate normality using the SOLAR software package [16]. To assess the effect of accounting for family structure on the test of association, we ran the models as described both accounting for and ignoring the g term. Constraining the g term to zero reduces the model to a conventional regression model.

One thousand data sets were generated for each combination of allele frequency, effect size and residual heritability. Power was defined as the proportion of data sets in which a significant association was detected at an alpha level of 0.05 by the likelihood ratio test. Simulations that provided statistically significant associations but whose effect estimates were in the opposite direction of the simulated effect were not considered true positives. Type 1 error rates were calculated as the proportion of times that a significant effect was observed at alpha level of 0.05 when no effect was simulated (i.e., when the genotypic means were simulated to be equal).

Results

Effect of Accounting for Family Structure on Point Estimates

Table 1 shows the estimated odds ratios obtained from analysis of the simulated data sets for discrete traits when accounting for and then ignoring family structure. The table shows the mean odds ratios recovered from the 5,000 simulated datasets (1,000 each for 5 allele frequencies) for both the sib pair and extended family designs. Also included is an empirical 95% confidence interval for each recovered odds ratio, that is, 95% of the simulations returned point estimates within the corresponding interval. The results shown are for a discrete trait with a residual heritability of 0.40. This table reveals that the simulated effect sizes were recovered very well in the analysis and that ignoring family structure does not bias the point estimate of the effect size under the assumption of random ascertainment. Additional simulations were carried out for other pedigree structures and across a range of heritabilities, and similar results were obtained, i.e., accurate recovery of the simulated SNP effect and no bias from ignoring family structure on effect size estimate (data not shown).

Intuition would suggest that ignoring family structure in the analysis should lead to wider confidence intervals around the point estimates, which is a statement of the precision of the estimate. Table 1 does not contradict this. The confidence intervals given in table 1, which are slightly smaller when accounting for family structure,

Table 2. Power of association tests to detect SNPs of moderate effects (odds ratio 1.2–2.0) when accounting for and ignoring family structure for a trait with residual heritability of 0.40. The allele frequency represents the frequency of the low risk allele

Allele frequency	Family structure	Odds ratio									
		1.2		1.4		1.6		1.8		2.0	
		ignore family	account for family	ignore family	account for family	ignore family	account for family	ignore family	account for family	ignore family	account for family
0.1	sib pair	0.11	0.10	0.30	0.27	0.51	0.49	0.65	0.63	0.80	0.79
	large sibship	0.13	0.11	0.30	0.27	0.49	0.46	0.69	0.64	0.77	0.73
	nuclear family	0.12	0.10	0.31	0.28	0.50	0.46	0.66	0.62	0.78	0.75
	extended family	0.15	0.10	0.30	0.26	0.48	0.43	0.66	0.60	0.75	0.72
	AFDS	0.13	0.10	0.33	0.30	0.48	0.45	0.65	0.62	0.78	0.77
0.2	sib pair	0.19	0.16	0.50	0.48	0.75	0.74	0.89	0.88	0.96	0.95
	large sibship	0.20	0.18	0.47	0.43	0.71	0.68	0.89	0.89	0.95	0.95
	nuclear family	0.19	0.17	0.49	0.44	0.74	0.70	0.90	0.88	0.96	0.96
	extended family	0.20	0.16	0.48	0.44	0.73	0.68	0.85	0.84	0.94	0.94
	AFDS	0.20	0.17	0.46	0.43	0.74	0.72	0.90	0.89	0.94	0.95
0.3	sib pair	0.26	0.23	0.60	0.58	0.85	0.83	0.96	0.96	0.99	0.98
	large sibship	0.25	0.21	0.58	0.54	0.85	0.82	0.96	0.95	0.98	0.98
	nuclear family	0.24	0.20	0.59	0.53	0.86	0.84	0.95	0.94	0.99	0.98
	extended family	0.24	0.21	0.55	0.52	0.81	0.79	0.94	0.94	0.99	0.99
	AFDS	0.25	0.24	0.62	0.57	0.85	0.83	0.94	0.94	0.99	0.99
0.4	sib pair	0.24	0.22	0.64	0.62	0.90	0.90	0.98	0.98	1.00	1.00
	large sibship	0.28	0.25	0.62	0.59	0.89	0.87	0.98	0.97	1.00	1.00
	nuclear family	0.25	0.22	0.67	0.63	0.87	0.86	0.98	0.98	1.00	1.00
	extended family	0.27	0.23	0.64	0.58	0.88	0.86	0.96	0.96	0.99	0.99
	AFDS	0.26	0.23	0.62	0.60	0.89	0.87	0.98	0.98	0.99	0.99
0.5	sib pair	0.26	0.25	0.68	0.67	0.91	0.90	0.99	0.99	1.00	1.00
	large sibship	0.26	0.23	0.66	0.62	0.90	0.88	0.98	0.98	1.00	1.00
	nuclear family	0.26	0.23	0.67	0.64	0.90	0.88	0.98	0.98	1.00	1.00
	extended family	0.28	0.24	0.64	0.61	0.88	0.87	0.98	0.98	1.00	0.99
	AFDS	0.27	0.23	0.67	0.64	0.90	0.89	0.98	0.97	1.00	1.00

pertain to the accuracy of the point estimate, not its precision. That is, on average, ignoring family structure does not bias the results, however there is slightly more scatter around the ‘true’ effect than there would be if family structure is considered.

Effect of Accounting for Family Structure on Power

The power to detect SNP effects is shown in table 2 for analyses that both account for and ignore family structure. A discrete trait with a residual heritability of 0.40 was simulated for effect sizes ranging from small (odds ratio = 1.2) to modest (odds ratio = 2.0). These analyses indicate that the power of the association test is relatively robust to the inclusion or exclusion of information on family structure. In every case, ignoring family structure gave virtually identical or slightly higher power than ac-

counting for family structure. However, in no case did the difference in power achieved by ignoring family structure exceed 6%.

Figure 2 illustrates the impact of accounting for family structure for differing levels of residual heritability. Models that accounted for and ignored family structure were run on the same set of simulated data and plotted for a common SNP with a minor allele frequency of 0.40 and odds ratio of 1.2. At low residual heritabilities (e.g., 10%, as shown in top panel), accounting for family structure makes very little difference and the p values for the two analytic approaches are virtually identical. At high trait heritabilities (e.g., 90%, as shown in bottom panel), the power to detect SNP effects is marginally higher when family structure is ignored, as shown by the fact that most of the data points fall above the diagonal line, indicating

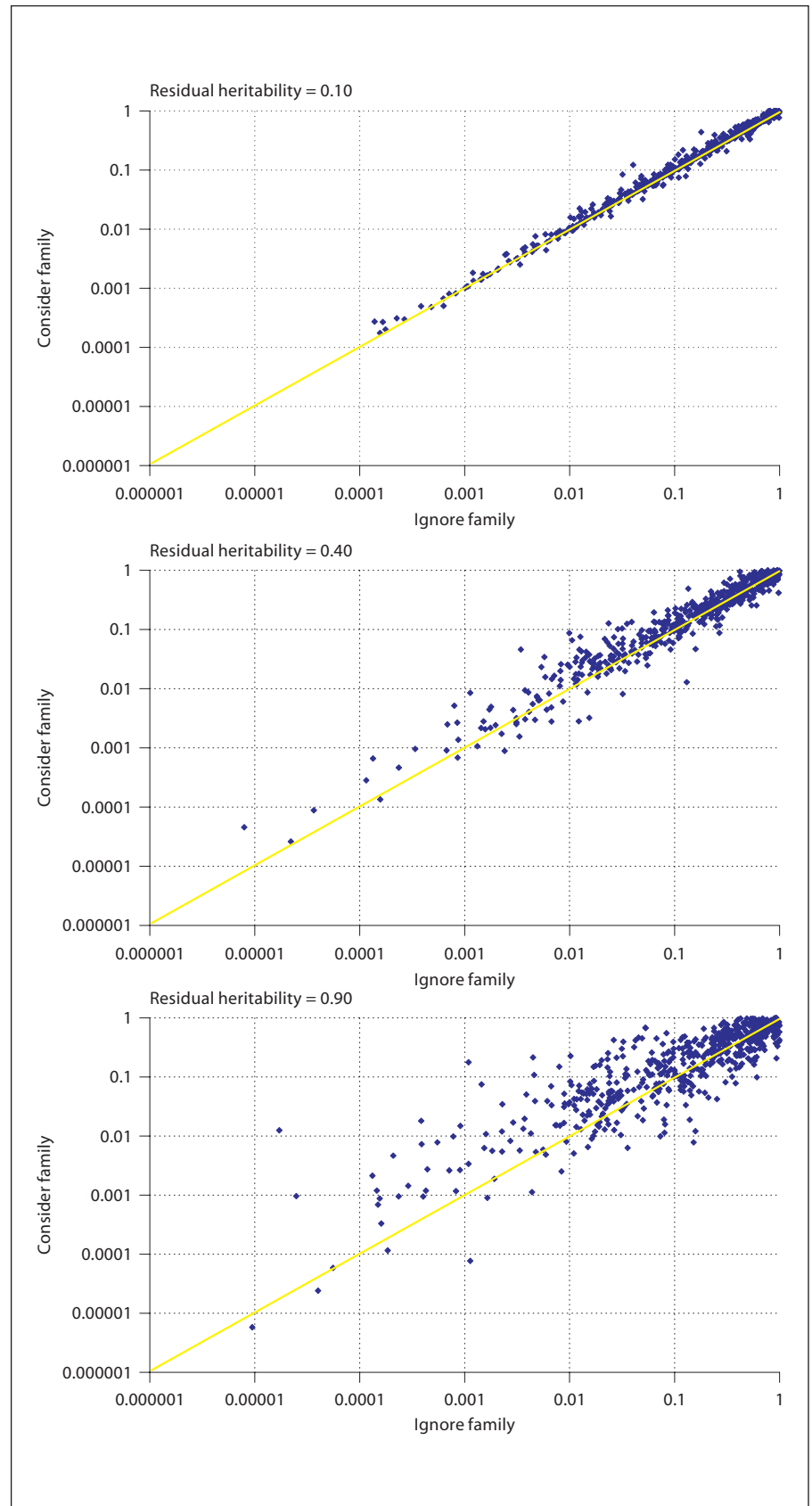


Fig. 2. Comparison of p values obtained considering and ignoring family structure for a common variant with a small effect (allele frequency = 0.40, odds ratio = 1.2) in the AFDS.

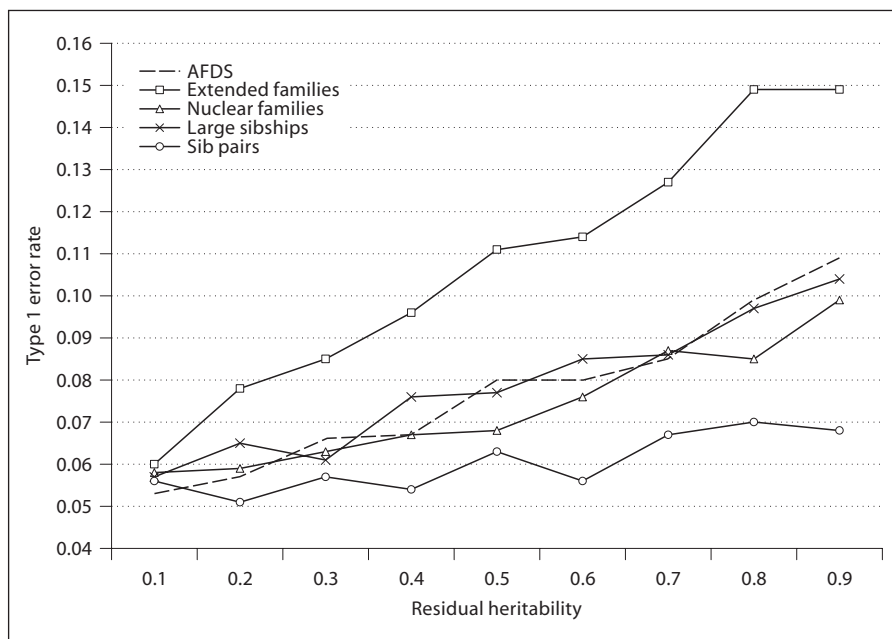


Fig. 3. Type 1 error rates by residual heritability for discrete traits for sib pairs, large sibships, nuclear families, extended families, and AFDS pedigrees when family structure is ignored.

that the significance of the association is inflated (i.e. the p value returned by ignoring relatedness is smaller than the one that accounts for family structure). When an alpha level is specified, most simulated datasets are either statistically significant by both models or not statistically significant by either model so that interpretation of the statistical significance of results is not appreciably affected by ignoring or considering family structure.

Effect on Type 1 Error

Type 1 error rates corresponding to the detection of a statistically significant association between SNP genotype and phenotype under the null hypothesis (i.e., no association) are shown in figure 3. The probabilities of falsely detecting an association when none was simulated are plotted against residual trait heritability for the simulated discrete trait when relatedness is ignored. Across all sampling designs, type 1 error rates rise as the residual heritability increases. Type 1 error rates also rise as the individuals in the sample are more related. For example, as indicated in figure 3, the sib pair design generally has the lowest type 1 error rates and the extended family design has the highest. Type 1 error rates for the AFDS are most similar to large sibship or nuclear family designs. For traits with levels of residual heritability that are common for many complex diseases, the type 1 error rate rises from the nominal 0.05 to 0.11 depending on sampling design. When family structure is appropriately modeled

in the analysis, the type 1 error rates for each sampling design are at the nominal 0.05 alpha level, as expected (data not shown).

Discussion

The objective of our study was to empirically contrast the performance of association analyses of family based designs that account for and ignore family structure in assessment of the phenotype-genotype association. This work is applicable both to tests of association among related individuals and transmission based tests of association in the presence of linkage that utilize a regression based framework, e.g. QTDT. This is an important area of inquiry because many currently available tools used to evaluate genome wide association data via regression based analyses ignore relatedness among phenotyped individuals in family based studies. The three major results from our analyses are that ignoring family structure: (1) has little impact on power to detect SNP effects when a consistent alpha level is used; (2) does not bias the estimation of the genotype odds ratio, and (3) increases type 1 error as a function of the residual heritability of the trait and the degree of relatedness of individuals in the sample. These three conclusions also hold for analysis of quantitative traits: that is, ignoring family structure in the analysis has little effect on power, does not bias the point es-

timate, and increases the type 1 error rate. However, the increase in type 1 error rate is marginally greater for quantitative traits as some information is necessarily lost during the assignment of affection status (data not shown).

Our analyses assumed the SNP under investigation is causative. If instead the SNP is in linkage disequilibrium with the causative variant, power would be reduced but our overall conclusions remain valid. Newman et al. highlighted the elevation of type 1 error rates that can occur by failing to account for genealogical relationships in a very large, complex Hutterite pedigree having a limited number of founders [17]. We considered a much simpler pedigree structure for our analyses since our intent was to generalize our results to pedigrees observed in more typical family studies.

Our results indicate that estimates of the effect of the SNP are not biased by ignoring family structure in the analysis. These results are perhaps predictable since the residual heritability, estimated using the family structure, should be uncorrelated with the SNP effect.

Equivalent power when ignoring family structure is maintained only at the expense of an increase in type 1 error. Because of the reciprocal relationship between power and type 1 error, maintaining a fixed type 1 error rate of 0.05 (as opposed to a pre-set alpha level of 0.05) across the family-based samples would result in a decrease in power. For example, using an adjusted alpha level to maintain a type 1 error rate of 0.05 for a trait with residual heritability of 0.40 would result in a power of 0.84 for a sib pair sample and 0.73 for the extended family design for an allele with a frequency of 0.3 having an effect of 1.6 odds ratio.

The elevation of type 1 error rates was related to increased residual heritability of the trait and greater relatedness among individuals in the sample. This relationship is expected since the degree of residual heritability is a function of the phenotypic correlations among related individuals, and with increasing heritability there is higher phenotypic covariance, and less 'independent' phenotypic information. However, it does not follow that for a given heritability, the degree of elevation of type 1 error will be consistent across all studies, since the phenotypic covariance also depends on the relatedness among individuals in the sample. One would expect type 1 error rates to increase less dramatically in samples that have relatively few relationship pairs since in such pedigrees the degree of correlation of phenotypic data would be lower. Indeed, our simulations of a sib pair study design led to smaller elevation of type 1 error rates than that

observed in larger pedigree structures. Thus, the degree to which the type 1 error rate will be elevated by ignoring family structure is a function not only of trait heritability, but also of pedigree size and relatedness among pedigree members. The type 1 error rates are given for four sampling designs as well as for an ongoing family study.

Generalized least squares estimates of both the parameter of interest and its variance are functions of residual heritability and relatedness among family members. We have shown empirically that the point estimate is unbiased. It follows that since statistical significance tends to be inflated, the variance of the estimate generally is underestimated when family structure is ignored. If the analysis is part of a large screening effort, then ignoring family structure will tend to screen SNPs into further consideration rather than missing a true association. If the initial screening is to be followed by a more comprehensive analysis, ignoring family structure may be considered for efficiency in an initial stage of high throughput screening analyses. Thus the more computationally intensive analyses that incorporate family structure need be applied only for the subset of markers identified in the first stage. This approach would efficiently identify the evidence for association within a dataset and properly estimate parameters for those markers of most interest.

As figure 3 demonstrates, there are occasions when ignoring family structure may return a less significant result than when accounting for family structure. This can occur because of movement of the point estimate towards the null or an overestimation of the variance based on sample variation and may lead to a false negative, i.e. a true association may be missed. If this is a concern, a less stringent alpha level can be used in stage 1 which would require more SNPs to be followed up in stage 2 but would also decrease the possibility of a false negative.

This paper empirically estimates the effects of ignoring familial relatedness in the analysis of genetic association study. This information should prove helpful to those designing analytic strategies for large scale genetic association studies.

Acknowledgements

This work was supported by NIH research grants HL69313, U01 HL72515, AR43351, DK54261, AR046838 and U01 HL084756.

References

- 1 Risch NJ: Searching for genetic determinants in the new millennium. *Nature* 2000; 405:847–856.
- 2 Daly MJ, Altshuler D: Partners in crime. *Nat Genet* 2005;37:337–338.
- 3 Middleton FA, et al: Complete maternal uniparental isodisomy of chromosome 4 in a subject with major depressive disorder detected by high density SNP genotyping arrays. *Am J Med Genet B Neuropsychiatr Genet* 2006;141:28–32.
- 4 Ozaki K, et al: Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002;32:650–654.
- 5 Klein RJ, et al: Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385–389.
- 6 Marchini J, Donnelly P, Cardon LR: Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 2005;37:413–417.
- 7 Ionita I, Man M: Optimal two-stage strategy for detecting interacting genes in complex diseases. *BMC Genet* 2006;7:39.
- 8 Kleinbaum DG: Applied regression analysis and other multivariable methods, ed 3. Pacific Grove, Duxbury Press, 1998.
- 9 Allison DB, Neale MC: Joint tests of linkage and association for quantitative traits. *Theor Popul Biol*, 2001;60:239–251.
- 10 Allison DB: Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 1997;60:676–690.
- 11 Abecasis GR, Cookson WO, Cardon LR: Pedigree tests of transmission disequilibrium. *Eur J Hum Genet* 2000;8:545–551.
- 12 Goddard ME: A mixed model for analyses of data on multiple genetic markers. *Theor Appl Genet* 1992;83:878–886.
- 13 Bull SB, et al: Evaluation of genetic and environmental effects using GEE and APM methods. *Genet Epidemiol* 1995;12:729–734.
- 14 Hopper JL, Mathews JD: Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet* 1982;46:373–383.
- 15 Hsueh WC, et al: Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care* 2000;23:595–601.
- 16 Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;62:1198–1211.
- 17 Newman DL, et al: The importance of genealogy in determining genetic associations with complex traits. *Am J Hum Genet* 2001; 69:1146–1148.