Biostatistics (2003), **4**, 2, *pp.* 195–206
Printed in Great Britain

A multivariate family-based association test using generalized estimating equations: FBAT-GEE

CHRISTOPH LANGE*
Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA
clange@hsph.harvard.edu

EDWIN K. SILVERMAN
Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA

XIN XU
Program of Population Genetics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA

SCOTT T. WEISS
Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA

NAN M. LAIRD
Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA

SUMMARY

In this paper we propose a multivariate extension of family-based association tests based on generalized estimating equations. The test can be applied to multiple phenotypes and to phenotypic data obtained in longitudinal studies without making any distributional assumptions for the phenotypic observations. Methods for handling missing phenotypic information are discussed. Further, we compare the power of the multivariate test with permutation tests and with using separate tests for each outcome which are adjusted for multiple testing. Application of the proposed test to an asthma study illustrates the power of the approach.

Keywords: Family-based association tests (FBATs); GEE; Genetic association tests; Multivariate phenotypes.

1. INTRODUCTION

Studies of association between disease outcomes and genetic markers often use samples of diseased subjects along with their parents or other family members. Such samples can be used to construct family-based association tests (FBATs) (Thomson, 1995; Zhao, 2000; Laird *et al.*, 2001). We use the term FBAT very generally to denote an association test which uses genetic data on family members to compute the distribution of a suitable test statistic under the null hypothesis, conditioning on the phenotypes. FBATs

*To whom correspondence should be addressed

provide simple and powerful tests to detect linkage between a marker and a disease susceptibility locus, in the presence of linkage disequilibrium between the two loci. The best known FBAT, the TDT proposed by Spielman *et al.* (1993), was developed for dichotomous disease traits under the assumption of having known genotypes for both parents and at least one affected offspring. Other tests have been suggested for quantitative phenotypes: Rabinowitz (1997); Allison *et al.* (1999) and Abecasis *et al.* (2000). Approaches for multiple phenotypes or longitudinal measures of phenotypes have not been developed.

Many genetic studies collect data on a set of related phenotypes that are potentially associated with the marker loci under study. One approach is to test all phenotypes individually and to adjust the p-value for multiple testing by the bonferroni correction. In practice, correcting the p-values by Bonferroni is conservative and many genetic studies may fail to detect true associations. Modifications of the Bonferroni correction have been proposed: for example, Hochberg (1988). Another alternative is to use permutation tests which permute the observed data under the null-hypothesis.

In the univariate phenotype setting, most FBAT statistics are based on the general likelihood-score test approach discussed in Lazzeroni and Lange (2001); Schaid and Sommer (1996) and Laird *et al.* (2000). Here we propose a natural extension of FBATs to multivariate phenotypes based on GEE scores (Liang and Zeger, 1986). Without making any assumptions about the distributions of the phenotypes, the multivariate FBAT test allows us to test the null hypothesis that the marker locus is not linked to any genetic locus that has an influence on the selected phenotypes. The proposed methodology is illustrated by an application to an asthma genetics study. While multiple FBAT-testing with various corrections failed to find a significant association, the multivariate test gives a significant result. Simulation experiments involving quantitative traits show that the multivariate FBAT clearly outperforms permutation tests and individual FBATs with corrections for multiple testing.

2. EXTENSION OF FBAT TO MULTIVARIATE DATA

In this section we will introduce an extension of FBATs for multivariate phenotypes which we subsequently refer to as *FBAT-GEE*. Since we compute the distribution of the test statistic conditional on the observed phenotypes, FBAT-GEE is a valid multivariate test that does not require any distributional assumption for the phenotypes. Hence FBAT-GEE can be applied directly to multiple phenotypic observations of arbitrary types: for example, dichotomous outcome variables, counts, continuous variables and to combinations of different types of variables.

For simplicity of exposition, we assume that we observe n independent families consisting of parents and one offspring, but the method extends easily to the more general case where we observe siblings and/or one or more parent is missing using the approach described in Rabinowitz and Laird (2000). We test the null-hypothesis that a marker locus is not linked to any disease-susceptibility locus for any of m selected phenotypes. The proposed idea can easily be adapted to situations when more than one linked/unlinked loci are included. However, to outline the idea it is sufficient to discuss this simpler one-locus scenario.

Assume first that we observe one bi-allelic marker locus with alleles A and B, with x_i counting the number of transmitted A alleles in the offspring of the ith family, i.e. $x_i = 0, 1, 2$. Other choices of genotype coding are possible (Schaid and Sommer, 1994). The parental genotypes for the ith family are denoted by p_{i1} and p_{i2}. Suppose there is a single phenotype and let the phenotypic observation for the offspring in the ith family be y_i. For such setups, Lazzeroni and Lange (2001); Schaid and Sommer (1996) and Laird *et al.* (2000) pointed out that univariate FBATs can be derived quite generally from likelihood scores where the likelihood function models the phenotype Y_i given the marker score x_i. Under the assumption that Y_i given x_i can be modelled by a generalized linear model with distributions from the exponential family—for example, Bernoulli for binary traits, normal distribution for continuous traits—conditioning on the sufficient statistic for any nuisance parameter under the null-hypothesis, Lunetta *et*

al. (2000) showed that the likelihood score is given by the statistic $\sum_{i=1}^{n} t_i x_i$ where t_i is an appropriate coding of the phenotype y_i. When trios with a single affected offspring are given, setting $t_i = 1$ yields the standard TDT proposed by Spielman *et al.* (1993). When also unaffected offspring are sampled, then coding affected offspring by $y_i = 1$ and unaffected by $y_i = 0$ and setting $t_i = y_i - K$ gives the FBAT proposed by Whittaker and Lewis (1998) where K is the population prevalence of the disease. For quantitative phenotypes y_i, we define $t_i = y_i - \overline{y}$ where $\overline{y}$ is the sample mean and obtain so the quantitative FBAT proposed by Rabinowitz (1997). Notice that setting $t_i = 0$ means that the ith offspring does not contribute to the test statistic (see below).

The statistic $\sum_{i=1}^{n} t_i x_i$ can then directly be utilized to construct a FBAT χ^2, ie

$$\chi^{\mathbf{2}} = \frac{(S - E(S))^2}{V_S} \quad \sim \quad \chi_1^2 \tag{1}$$

with the statistic $S = \sum_{i=1}^{n} t_i x_i$. Its expected value is $E(S) = \sum_{i=1}^{n} t_i E(x_i \,|p_{i1}, p_{i2})$ and its variance is $V_S = \sum_{i=1}^{n} t_i^2 \,\mathrm{Var}(x_i \,|p_{i1}, p_{i2})$. The mean and the variance of the marker, $E(x_i \,|p_{i1}, p_{i2})$ and $\mathrm{Var}(x_i \,|p_{i1}, p_{i2})$, are computed under the null-hypothesis conditioned on the parental genotypes.

When we observe m phenotypes instead of only one phenotype per offspring, the univariate χ^2 can be used to test each phenotype individually, but it cannot test all phenotypes simultaneously. However, its generalization to multivariate data is straightforward. We denote the m-dimensional vector containing all the phenotypic information for the ith offspring by $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})$ and use the multivariate score based on the generalized estimating equation (GEE) approach (Liang and Zeger, 1986; Heyde, 1997). The GEE score is then given by the statistic $\mathbf{S} = \sum_{i=1}^{n} x_i \; \mathbf{\Delta}_i \,\mathrm{Var}(\mathbf{t}_i)^{-1}\mathbf{t}_i$, where $\mathrm{Var}(\mathbf{t}_i)$ is the 'working' variance matrix for the phenotypes of the ith individual, $\mathbf{\Delta}_i$ is a diagonal matrix that depends on the model assumptions of the underlying GEE model and $\mathbf{t}_i$ is the m-dimensional vector containing the coded trait information: for example, for continuous phenotypes $\mathbf{t}_i = \mathbf{y}_i - \overline{\mathbf{y}}$ where $\overline{\mathbf{y}}$ is the vector of observed sample means. With no missing phenotypic data and no covariates, the variance matrix of $\mathbf{t}_i$ and $\mathbf{\Delta}_i$ are identical for all subjects under H_0, ie $\mathrm{Var}(\mathbf{t}) \equiv \mathrm{Var}(\mathbf{t}_i)$ and $\mathbf{\Delta} \equiv \mathbf{\Delta}_i$. Thus, they vanish when the score test is constructed under the null-hypothesis. Hence we define the m-dimensional centered score vector $\tilde{\mathbf{S}}$ by

$$\tilde{\mathbf{S}} = \sum_{i=1}^{n} \mathbf{t}_i \,(x_i - E\,(x_i \,|p_{i1}, p_{i2}))\,. \tag{2}$$

Further, note that $E\left(\tilde{\mathbf{S}}\right) = \mathbf{0}$. It is then easy to see that the $m \times m$ variance matrix of the vector $\tilde{\mathbf{S}}$ is given by

$$\mathbf{V}_{\tilde{\mathbf{S}}} = \mathrm{Var}\left(\tilde{\mathbf{S}}\right) = \sum_{i=1}^{n} \mathbf{t}_i \mathbf{t}_i^t \,\mathrm{Var}\,(x_i \,|p_{i1}, p_{i2})\,. \tag{3}$$

So the multivariate extension of the univariate FBAT can be defined by

$$\chi^{\mathbf{2}}_{\mathbf{FBAT-GEE}} = \tilde{\mathbf{S}}^t \; \mathbf{V}_{\tilde{\mathbf{S}}}^{-1} \; \tilde{\mathbf{S}}. \tag{4}$$

Asymptotic theory (Lange and Laird, 2002) therefore implies that the multivariate FBAT $\chi^{\mathbf{2}}_{\mathbf{FBAT-GEE}}$ is asymptotically χ^2-distributed,

$$\chi^{\mathbf{2}}_{\mathbf{FBAT-GEE}} \sim \chi^2_k, \tag{5}$$

where the degrees of freedom $k \leqslant m$ are given by the rank of the variance matrix $\mathbf{V_S}$, i.e. $k = rank\,(\mathbf{V_S})$. Unless there are linear dependences between the m phenotypes, $k = m$. In general, since all phenotypes

are incorporated simultaneously in the computation of $\chi^2_{\mathbf{FBAT-GEE}}$, the FBAT-GEE test uses more information than individual FBATs and should therefore be more powerful.

The interpretation of a significant FBAT-GEE test result depends on assumptions about the underlying genetic model. When one assumes that all phenotypes tested by FBAT-GEE are linked to an unobserved trait/genetic mechanism which is influenced by the tested gene, a significant FBAT-GEE can be understood as a significant result for the underlying trait/mechanism, which is expressed by the selected phenotypes.

When one assumes no shared genetic mechanism and the main purpose of FBAT-GEE is to reduce the number of comparisons by testing jointly all phenotypes in one symptom group, a significant FBAT-GEE result implies that one has to test all phenotypes of a significant group individually by FBATs to detect the associated traits. Since the individual univariate FBATs are only applied when FBAT-GEE was significant, the significance level of FBAT-GEE is maintained and is also the overall significance level. For example, assume we want to test 30 phenotypes for association at an overall significance level of α. The traits can be grouped into 5 subgroups. Then using individual FBATs, the adjusted significance level is $\alpha/30$ (Bonferroni correction). However, when we first test the five groups by FBAT-GEE the adjusted significance is $\alpha/5$. Assuming that one group shows significance, we test all six phenotypes within this group individually, each at an adjusted significance level of $\alpha/6$. Thus, the overall significance level and the significance level in each step, FBAT-GEE and then univariate FBATs, is α.

When a multi-allelic marker locus is considered, the FBAT-GEE test extends naturally. The marker information $\mathbf{x}_i$ becomes a p-dimensional vector giving marker information for multiple alleles,

$$\tilde{\mathbf{S}} = \sum_{i=1}^{n} \mathbf{t}_i \otimes (\mathbf{x}_i - E\,(\mathbf{x}_i \,|p_{i1},\, p_{i2}))\,,$$

where $\otimes$ denotes the Kronecker product. Note that $\tilde{\mathbf{S}}$ is now a pm-dimensional vector. The variance of $\tilde{\mathbf{S}}$ is a $pm \times pm$ matrix given by

$$\mathrm{Var}\left(\tilde{\mathbf{S}}\right) = \sum_{i=1}^{n} \left\{\mathbf{t}_i \mathbf{t}_i^t\right\} \otimes \mathrm{Var}\,(\mathbf{x}_i \,|p_{i1},\, p_{i2})\,.$$

$E\,(\mathbf{x}_i \,|p_{i1},\, p_{i2})$ and $\mathrm{Var}\,(\mathbf{x}_i \,|p_{i1},\, p_{i2})$ for the multiallelic case are given in Horvath *et al.* (2001). The FBAT-GEE test for a multi-allelic locus can also be computed by (4) and has an asymptotic χ^2-distribution under the null-hypothesis where the degrees of freedom are given by the rank of $\mathrm{Var}(\tilde{\mathbf{S}})$. Since $\mathrm{Var}\,(\mathbf{x}_i \,|p_{i1},\, p_{i2})$ generally has rank $p-1$, $\mathrm{Var}(\tilde{\mathbf{S}})$ has rank $m(p-1)$ as long as all phenotypes are linearly independent. It is important to note that for alternative genotypic codings, eg for dominant or recessive diseases, $\mathrm{Var}\,(\mathbf{x}_i \,|p_{i1},\, p_{i2})$ may have rank p. Then $\mathrm{Var}(\tilde{\mathbf{S}})$ has rank mp.

The multivariate FBAT $\chi^2_{\mathbf{FBAT-GEE}}$ and its validity do not depend upon any distributional assumption for the phenotypes. This model-free character of $\chi^2_{\mathbf{FBAT-GEE}}$ has the great advantage that it allows us to test different trait types (e.g. binary phenotypes, continuous phenotypes, etc.) simultaneously without having to specify a statistical model which describes the dependence of the multivariate phenotypic vector $\mathbf{y}_i$ on the marker score x_i.

Note that $\chi^2_{\mathbf{FBAT-GEE}}$ as defined here is a valid test for the null-hypothesis of no linkage if siblings are used, because transmissions of genes from parents to different offspring are independent when there is no linkage. For the null-hypothesis of no association in the presence of linkage, transmissions are correlated (Martin *et al.*, 2000). However, $\chi^2_{\mathbf{FBAT-GEE}}$ can easily be modified to test for association using the empirical variance proposed by Lake *et al.* (2000). Generalization of FBAT-GEE for missing parental information and/or for missing phenotypic information are also straightforward. When parental genotypic information is missing, the conditional marker mean and variance under the null-hypothesis can be computed by using the theory of family-based association tests (Rabinowitz and Laird, 2000; Laird *et al.*, 2000).

When phenotypic information is missing, we have the choice between two approaches, imputing the missing phenotypes or using the GEE score of the 'observed' phenotypes. For example, the phenotypes can be imputed by the observed phenotypic mean, which is equivalent to setting $t_{ij} = 0$ when t_{ij} is defined by $t_{ij} = y_{ij} - \overline{y}_j$. Note that this effectively means that the ith subject contributes no information to the test statistic for the jth phenotype. A more sophisticated way would be to imputed the missing phenotypes by the EM-algorithm (Dempster *et al.* 1977), estimating the mean and variance matrix from the entire sample. However, no matter which imputation technique we elect they all suffer from the conceptional drawback of being computed under the null-hypothesis. The imputed phenotypes will therefore always provide evidence for the null-hypothesis and will make the test $\chi^2_{\mathbf{FBAT-GEE}}$ more conservative. On the other hand, applying imputation techniques under the alternative hypothesis would require making model assumptions about the distribution of the phenotypes.

Alternatively, in presence of missing phenotypic information, the 'observed' GEE score can be utilized. The GEE score is now given by $\mathbf{S} = \sum_{i=1}^{n} x_i \mathbf{I}_i \boldsymbol{\Delta}_i \, \mathrm{Var}(\mathbf{t}_i)^{-1} \mathbf{t}_i$ where $\mathbf{t}_i$ is the vector of observed traits. $\mathbf{I}_i$ is constructed by taking a $m \times m$ identity matrix and removing the columns that correspond to the missing phenotypic information. Since the matrices $\boldsymbol{\Delta}_i$ and $\mathrm{Var}(\mathbf{t}_i)$ are constructed by taking $\boldsymbol{\Delta}$ and $\mathrm{Var}(\mathbf{t})$ and removing the rows and columns that correspond to the missing observations, the dimensions of $\boldsymbol{\Delta}_i$ and $\mathrm{Var}(\mathbf{t}_i)$ can vary between families. The variance matrix $\mathrm{Var}(\mathbf{t}_i)$ and $\boldsymbol{\Delta}_i$ do therefore not vanish in the construction of score test statistic and have to be estimated when the GEE score is used for the computation of the test statistic. We will apply both approaches of handling missing phenotypes, imputation techniques and 'observed' GEE scores, to the asthma study.

Further, it is important to note that $\chi^2_{\mathbf{FBAT-GEE}}$ is invariant under any linear transformation. Since principal component analysis or canonical correlations are special cases of such linear transformations, they can only be useful to reduce the number of phenotypes tested. Simply transforming phenotypes will not have any effect either on the p-value or on the power.

3. Data analysis: childhood asthma management program

We applied the FBAT-GEE approach to a collection of parent/child trios in the Childhood Asthma Management Program (CAMP) Genetics Ancillary Study. The CAMP study is a clinical trial of 1041 asthmatic children who were randomized to three different asthma treatments (CAMP, 1999). Blood samples for DNA were collected from 696 complete parent/child trios from 640 pedigrees in the CAMP Ancillary Genetics Study. Baseline phenotype values, before randomization to treatment groups, were used in this analysis. Genotyping was performed at a polymorphism located at amino acid 16 of the Beta-2 Adrenergic Receptor (B2AR) gene. Previous case/control association studies have reported an association of the B2AR-16 polymorphism with various phenotypes including bronchodilator responsiveness, nocturnal asthma, and pulmonary function (the forced expiratory volume at one second of a forced expiratory maneuver, or FEV1) (Turki *et al.*, 1995; Martinez *et al.*, 1997; Summerhill *et al.*, 2000). Asthma is a clinical condition often associated with an atopic predisposition; we have selected three phenotypes related to the allergic response, including total eosinophil count, core number of positive allergy skin tests to common environmental antigens, and total serum immunoglobulin E levels (expressed as $\log_{10}$). Asthma is also associated with a variety of abnormalities in pulmonary function, including reductions in forced expiratory flow (such as the FEV1), increased tendency for bronchoconstriction in response to chemicals such as methacholine (expressed as the natural logarithm of the provocative concentration causing a 20% reduction in FEV1 or $\ln(PC20)$), and increased FEV1 following administration of bronchodilator medications. We included three pulmonary function phenotypes, FEV1 (expressed as percentage of predicted after bronchodilator treatment), $\ln(PC20)$ and bronchodilator responsiveness

Table 1. *Single phenotype analysis with separate FBATs at an overall significance level of* $\alpha = 0.05$*: significance level for the separate FBATs after Bonferroni-correction,* $\alpha_{\text{Bonferroni}} = 0.008$

Phenotype	N	$\overline{Y}$	$\sqrt{\text{Var}(Y)}$	S	$E(S)$	Var(S)	$\sqrt{FBAT}$	p-Value
FEV1	439	103	12.3	268.1	−92.634	20417	2.525	0.012
ln(PC20)	439	0.052	1.17	4.764	−6.018	201.685	0.759	0.448
BD increase	379	25.3	26.5	684.7	185.549	93217	1.635	0.102
Eosinophil count	422	517	459	−3468	−8817	27615752	1.018	0.309
log(IgE)	434	2.63	0.675	5.643	−0.927	70.329	0.783	0.433
Number of positive skin tests	393	5.56	4.27	68.34	43.616	2545	0.490	0.624

(expressed as change in FEV1 as a percentage of predicted FEV1 after bronchodilator treatment). A more comprehensive analysis of associations between B2AR-16 and baseline phenotypes will be presented separately.

We elected to test at significance level $\alpha = 0.05$ whether any of the phenotypes was associated with the biallelic marker locus (B2AR-16). Missing phenotypic data was handled either by imputation or by computing FBAT-GEE based on the 'observed' GEE scores where all traits were considered to be normally distributed and the phenotypic means and variance matrix were estimated by their empirical estimators based on the observed and imputed phenotypes (EM-algorithm). The phenotypes were imputed either by setting $t_{ij} = 0$ or by the EM-algorithm (Dempster *et al.* 1977). Of the 696 families, 439 had at least one heterozygous parents and thus contribute to the test statistic.

Initially, we tested for association between B2AR-16 with each phenotype separately, using the FBAT program http://www.biostat.harvard.edu/∼fbat/default.html). The p-values for family-based association testing of each phenotype separately are shown in Table 1. Although there was some evidence for association of B2AR-16 to FEV1, after Bonferroni correction none of the phenotypes were significant ($\alpha_{\text{Bonferroni}} = 0.008$).

Subsequently, we used FBAT-GEE, which is also implemented in the FBAT program, to test for association within the two groups of phenotypes, atopy-related and pulmonary function phenotypes (Table 2). The FBAT-GEE was computed three times: by setting $t_{ij} = 0$, imputing missing values by the EM-algorithm and based on observed GEE scores; since two groups were tested, the significance level was $\alpha_{\text{Bonferroni}} = 0.025$. For all approaches of handling missing phenotypic information at this level of significance, a significant association was detected between B2AR-16 and pulmonary function phenotypes; no significant association to atopy-related phenotypes was detected. The significant p-value (0.012) obtained by FBAT-GEE based on 'observed' GEE scores was slightly smaller than the p-value obtained by the imputation methods (0.021 and 0.018). Since the imputation methods used here assume that the null-hypothesis is true, it is likely that FBAT-GEE based on such imputed values is conservative. On the other hand, using the GEE scores does not require such assumption. Further studies are desirable to assess the optimal strategy.

Since the pulmonary function phenotypes showed a significant FBAT-GEE result, we test the phenotypes of this group individually. The overall significance level $\alpha = 0.05$ is maintained when the individual FBAT for the three phenotypes achieve jointly a significance level of $\alpha = 0.05$. We therefore use the Bonferroni correction again. The individual FBATs have to be significant at $\alpha = 0.05/3 = 0.016$. Thus, FEV1 reaches overall significance (Table 1). Because FEV1 is not highly correlated with other pulmonary function phenotypes, we conclude that there is good evidence for an effect of B2AR-16 only on FEV1, and no evidence for atopy phenotypes or other pulmonary function phenotypes.

Table 2. *Joint phenotype analysis with FBAT-GEE at significance level of* $\alpha = 0.025$. *Missing phenotypes were handled either by mean imputation, ie* $t_{ij} = 0$, *by imputation based on EM or by 'observed' GEE scores*

Method	Group	N	FBAT-GEE	df	p-value
Mean imputation	Pulmonary function phenotypes	439	9.73	3	0.021
	Atopy-related phenotypes	434	2.16	3	0.54
Imputation based on EM	Pulmonary function phenotypes	439	10.068	3	0.018
	Atopy-related phenotypes	434	1.916	3	0.59
GEE score	Pulmonary function phenotypes	439	10.95	3	0.012
	Atopy-related phenotypes	434	0.849	3	0.84

4. Power comparison between FBAT-GEE, permutation tests and multiple adjusted FBAT-testing

In this section we assess the power of the FBAT-GEE test by simulation experiments and compare it with permutation tests and with multiple FBAT-testing where the p-value is adjusted either by Bonferroni correction (BC) or by Hochberg (1988) correction (HC). The HC-method is a sequential version of the BC-method. The scenarios addressed in this simulation experiment are motivated by the asthma study. We assume that a sample of asthmatics and their parents are available; we observe the genotypes of all family members. For each asthmatic, two groups of three normally distributed phenotypes are measured. We want to test each group for a potential association with the marker locus.

We generate trios with a biallelic marker locus by drawing the parental genotypes p_1 and p_2 from the Bionomial distribution $Bi(2, p)$ where p is the allele frequency of the disease gene in the population. Using Mendelian transmissions, the genotype of the individual x_i is then simulated based on the parental genotypes. The phenotypic vector $\mathbf{Y}_i$ for each individual is obtained by generating samples from a multivariate normal distribution, i.e. $\mathbf{Y}_i \sim N((a_1 x_i, a_2 x_i, a_3 x_i), \mathbf{\Sigma})$, where a_j, $j = 1, 2, 3$ are the additive effects for the jth phenotype and $\mathbf{\Sigma} = \left(\sigma_{ij}^2\right)$ is the (3×3) variance matrix. The strength of an additive effect relative to the phenotypic variance is expressed by the *heritability* h_j^2 (Falconer and Mackay, 1997), which is is defined as the the proportion of phenotypic variation explained by the genetic variation, ie $h_j^2 = \mathrm{Var}\left(a_j X_i\right)/\mathrm{Var}\left(Y_{ij}\right)$. This equation can be solved for a_j and an analytical expression for a_j given h_j^2 can be obtained, $a_j = \sigma_{jj}\sqrt{h_j/2p\,(1-p)\left(1-h_j\right)}$.

In the asthma study we analyzed two groups of phenotypes, pulmonary function phenotypes and atopy-related phenotypes. The empirical correlation matrices of these groups are used in the simulation study as correlation matrices for the phenotypes; these are given in Table 3. For simplicity, we assume the variances $\sigma_{ii}^2 = 1$ and use the correlation matrices of those groups as variance matrices. We conduct the first simulation study under the assumption that all three phenotypes of the pulmonary function group are associated with the marker locus. We therefore set the heritability of all three phenotypes to 0.05, ie $h_1 = h_2 = h_3 = 0.05$, and $\mathbf{\Sigma}$ equal to the correlation matrix $\mathbf{\Sigma}_1$ of the pulmonary function phenotypes (Table 3). We generate $n = 300$ trios where each offspring has three phenotypes.

Then a number of tests are applied. First, we compute FBAT-GEE for the first phenotype. For a univariate phenotype, FBAT-GEE is identical to the continuous FBAT proposed by Rabinowitz (1997). Then we compute FBAT-GEE for the first two phenotypes and finally FBAT-GEE is calculated for all three phenotypes. All observed FBAT-GEE values are compared to the appropriate threshold values. Then the same sets of phenotypes are tested for association by testing the phenotypes individually by the univariate continuous FBAT (Rabinowitz, 1997) and comparing the observed test statistic values to

Table 3. *Power comparison: FBAT-GEE versus multiple testing by univariate FBATs and adjusting either by Hochberg correction (HC) or by Bonferroni correction (BC). The sample size is 300 and the significance level* $\alpha = 0.01$

Correlation matrix	Allele frequency p	N_{fam}	Test	$m = 1$	$m = 2$	$m = 3$
Scenario I: All phenotypes are associated with the marker locus				$h = 0.05$	$h = 0.05$	$h = 0.05$
$\Sigma_1 = \begin{pmatrix} 1.00 & -0.03 & 0.10 \\ -0.03 & 1.00 & -0.12 \\ 0.10 & -0.12 & 1.00 \end{pmatrix}$	0.3	199	FBAT-GEE	0.56	0.84	0.94
			Permutation	0.57	0.71	0.82
			FBAT & HC	0.56	0.73	0.81
			FBAT & BC	0.56	0.72	0.79
	0.1	98	FBAT-GEE	0.53	0.77	0.88
			Permutation	0.54	0.73	0.79
			FBAT & HC	0.53	0.68	0.75
			FBAT & BC	0.53	0.66	0.74
	0.05	54	FBAT-GEE	0.46	0.66	0.75
			Permutation	0.45	0.64	0.68
			FBAT & HC	0.46	0.60	0.65
			FBAT & BC	0.46	0.57	0.63
Scenario II: Only the first phenotype is associated with the marker locus				$h = 0.05$	$h = 0.00$	$h = 0.00$
$\Sigma_2 = \begin{pmatrix} 1.00 & 0.43 & 0.27 \\ 0.43 & 1.00 & 0.48 \\ 0.27 & 0.48 & 1.00 \end{pmatrix}$	0.3	199	FBAT-GEE	0.56	0.56	0.49
			Permutation	0.56	0.49	0.45
			FBAT & HC	0.56	0.47	0.41
			FBAT & BC	0.56	0.47	0.41
	0.1	98	FBAT-GEE	0.53	0.51	0.43
			Permutation	0.52	0.43	0.39
			FBAT & HC	0.53	0.43	0.37
			FBAT & BC	0.53	0.43	0.37
	0.05	54	FBAT-GEE	0.47	0.43	0.34
			Permutation	0.48	0.40	0.33
			FBAT & HC	0.47	0.36	0.30
			FBAT & BC	0.47	0.36	0.30

Estimated power and significance levels are shown based on 1 000 000 replicates. 'FBAT & BC' denotes the continuous FBAT by Rabinowitz (1997) adjusted by Bonferroni correction for multiple testing. 'FBAT & HC' denotes the continuous FBAT by Rabinowitz (1997) adjusted by Hochberg correction (1988) for multiple testing. The number of informative families is given by N_{fam}.

threshold values that have been adjusted either by HC or by BC. Further, a permutation test is computed where the phenotypic vectors are kept fixed and the trio's genotypes are randomly permuted. As for FBAT-GEE, the permutation test was computed for one, two and three phenotypes. This procedure is repeated for 10^6 replicates. The allele frequency is assumed to be $p = 0.3, 0.1, 0.05$. The significance level is $\alpha = 0.01$. The results are given in Table 3. Table 3 shows the estimated power for all FBAT-GEEs,

Table 4. *Estimated significance levels: FBAT-GEE versus multiple testing by univariate FBATs and adjusting either by HC or by BC. The sample size is 300 and the significance level* $\alpha = 0.01$. *Scenario: no phenotype is associated with the marker locus*

Allele frequency p	N_{fam}	Test	$m = 1$ $h = 0.00$	$m = 2$ $h = 0.00$	$m = 3$ $h = 0.00$
0.3	199	FBAT-GEE	0.010	0.0099	0.010
		Permutation	0.012	0.013	0.011
		FBAT & HC	0.011	0.010	0.010
		FBAT & BC	0.010	0.0098	0.0097
0.1	98	FBAT-GEE	0.0093	0.0088	0.0084
		Permutation	0.008	0.009	0.009
		FBAT & HC	0.011	0.008	0.008
		FBAT & BC	0.0093	0.0087	0.0084
0.05	54	FBAT-GEE	0.0087	0.0078	0.0070
		Permutation	0.008	0.010	0.012
		FBAT & HC	0.008	0.008	0.008
		FBAT & BC	0.0087	0.0079	0.0075

Estimated power and significance levels are shown based on 1 000 000 replicates. 'FBAT & BC' denotes the continuous FBAT by Rabinowitz (1997) adjusted by Bonferroni correction for multiple testing. 'FBAT & HC' denotes the continuous FBAT by Rabinowitz (1997) adjusted by Hochberg correction (1988) for multiple testing. The number of informative families is given by N_{fam}.

permutation tests and the individual FBATs adjusted either by HC or BC. Further, we estimated the true significance levels for $\alpha = 0.01$ (Table 4).

The same simulation experiment is repeated for the second group of phenotypes of the asthma study. For the atopy-related phenotypes, we assume that only the first phenotype is associated with marker locus, while the other two phenotypes are not, i.e. $h_1 = 0.05$, $h_2 = h_3 = 0$. The correlation matrix is now given by $\mathbf{\Sigma}_2$ (Table 3). The estimated power for this scenario is shown in Table 3.

For both scenarios, Table 3 shows that FBAT-GEE is preferable to permutation tests and testing by individual FBATs and adjusting either by HC or by BC. When all phenotypes are associated with the marker locus, the power of FBAT-GEE increases much faster than that of the other tests. On the other hand, when only the first phenotype is associated with the locus, the power of FBAT-GEE drops more slowly than the power for the other tests as further unassociated phenotypes are added. The overall ranking of the tests which can be observed in each simulation experiment is FBAT-GEE, permutations tests, individually testing with HC and individually testing with BC. While FBAT-GEE performs considerably better than permutation tests, permutation tests clearly outperform individually testing. However, HC does only a little bit better than BC.

The empirical significance levels for multiple testing with HC or BC are very close to those of FBAT-GEE. This suggests that the higher power of FBAT-GEE is due to the fact that FBAT-GEE considers all phenotypes simultaneously rather than uses only the information of one phenotype at a time. This property makes FBAT-GEE a promising test that has many applications to multivariate data, repeated measures and longitudinal data. Table 3 also shows that FBAT-GEE and individual FBATs adjusted by corrections become more and more conservative for smaller values of p. In a total population sample a smaller allele frequency p is equivalent to a smaller number of informative families. Further, we observe in

Table 4 that the estimated significance level for FBAT-GEE decreases when the number of phenotypes m increases. This observation indicates that the convergence speed of FBAT-GEE towards its asymptotic distribution depends on the number of phenotypes, ie the distribution of FBAT-GEE converges more slowly for increasing m.

The permutation test is slightly anti-conservative. In addition, we note that applying the permutation test is much more complex with multiple siblings, missing parents and/or phenotypes, etc or when linkage is present.

5. DISCUSSION

In this paper we have presented a multivariate approach to transmission disequilibrium tests that allows us to test several phenotypes simultaneously for linkage with one marker locus. Although different trait types can be tested simultaneously, no assumptions about their distributions have to be made. Further, various extensions of the FBAT-GEE approach are possible. Since we only have to know the marker mean and variance under the null-hypothesis for the computation of the test statistic, the FBAT-GEE approach can easily be adapted to scenarios with multiple offspring per family and missing parental information, and testing for linkage disequilibrium under the assumption of linkage. When phenotypic data are missing the FBAT-GEE approach can easily be extended either by imputation methods or by 'observed' GEE scores.

We have compared the power of FBAT-GEE, permutation tests and multiple testing with HC and BC by simulation experiments for scenarios derived from the asthma data set. Under these scenarios FBAT-GEE proved to be substantially more powerful than the other tests. We also repeated these simulations for combinations of different trait types: for example, counts, binary outcomes and percentages. Thereby we used the multivariate normal distribution to generate random numbers and then 'discreteized' the outcomes. The results of these simulation studies suggest that the superiority of the FBAT-GEE over the other tests decreases with the discreteness of the phenotypes. For example, while FBAT-GEE is clearly more powerful for count and percentage variables, this advantage becomes less relevant for binary outcome variables. Further simulation studies also supported the assumption that 'observed' GEE scores are a more powerful way of handling missing phenotypic information than imputation methods.

We have applied this novel approach for inclusion of multiple phenotypes to a relevant example in asthma genetics. For many genetic epidemiological studies, comprehensive phenotypic data is collected. However, use of multiple phenotypes in independent statistical tests requires correction for the multiple testing performed. Our proposed method provides a more efficient use of multiple phenotypes that does not require costly corrections for multiple comparisons. Further study will be required to determine the optimal approach to group phenotypes for such analyses. So far, our experience with real data sets is that FBAT-GEE is more powerful than individual FBAT testing when the phenotypes can be grouped into symptom groups with up to eight or nine phenotypes in each group: for example, here pulmonary function phenotypes and atopy-related phenotypes. Then testing each group individually and adjusting for multiple testing seems to be far more powerful than testing each phenotype individually (Demeo et al (in preparation)). A GEE approach for population data is discussed in Lange and Whittaker (2002).

ACKNOWLEDGEMENTS

We would like to thank Dr John C. Whittaker for his helpful comments on an earlier version of the paper. This research was supported in parts by grant MH59532 and in parts by grant HL66383 of The National Institutes of Health. The Childhood Asthma Management Program (CAMP) is supported by contracts from the NHLBI, including N01 HR16040. We acknowledge the CAMP investigators and research team for assistance in collection of the CAMP Genetics Ancillary Study data. Comments from two referees and an associate editor were very helpful in preparing this version of the paper.

REFERENCES

ABECASIS, G. R., CARDON, L. R. AND COOKSON, W. O. C. (2000). A general test of association for quantitative traits in nuclear families. *American Journal of Human Genetics* **66**, 279–292.

ALLISON, D. B. (1997). Transmission-disequilibrium tests for quantitive traits. *American Journal of Human Genetics* **60**, 676–690.

ALLISON, D. B., HEO, M., KAPLAN, N. AND MARTIN, E. R. (1999). Sibling-based tests of linkage and assoication for quantitive traits. *American Journal of Human Genetics* **64**, 1754–1764.

COX, D. R. AND HINKLEY, D. V. (1974). *Theoretical statistics*. London: Chapman and Hall.

CHILDHOOD ASTHMA MANAGEMENT PROGRAM RESEARCH GROUP (1999). The childhood asthma management program (camp): design, rationale, and methods. *Control Clinical Trials* **20**, 91–120.

DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statisitical Society* **39**, 1–38.

FALCONER, D. S. AND MACKAY, T. F. C. (1997). *Introdction to Quantitative Genetics*. Longman.

HEYDE, C. (1997). *Quasi-likelihood and its application*, Springer Series in Statistics. Berlin: Springer.

HOCHBERG, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800–802.

HORVATH, S. AND LAIRD, N. M. (1998). A discordant-sibship test for disequilibrium and linkage: no need for parental data. *American Journal of Human Genetics* **63**, 1886–1897.

HORVATH, S., XU, X. AND LAIRD, N. M. (2001). The family based association test method: strategies for studying general genotype-phenotype associations. *European Journal of Human Genetics* **9**, 301–306.

LAIRD, N. M., HORVATH, S. AND XU, X. (2000). Implementing a unified approach to family based tests of associaton. *Genetic Epidemiology* **19(Suppl 1)**, S36–S42.

LAKE, S. L., BLACKER, D. AND LAIRD, N. M. (2000). Family-based tests of association in the presence of linkage. *American Journal of Human Genetics* **67**, 1515–1525.

LANGE, C. AND LAIRD, N. M. (2002). On a general class of conditional tests for family-based association studies in genetics: the asymptotic distribution, the conditional power and optimality considerations. *Genetic Epidemiology* **23**, 165–180.

LANGE, C. AND WHITTAKER, J. C. (2001). A generalized estimating equation approach to mapping of quantitative trait loci (qtl). *Genetics* **159**, 1325–1337.

LANGE, C. AND WHITTAKER, J. C. (2002). Generalized estimating equations: a hybrid approach for mean parameters in multivariate regression models. *Statistical Modelling* in press.

LAZZERONI, L. C. AND LANGE, K. (2001). A conditional inference framework for extending the transmission/disequilibrium test. *Human Heredity* **48**, 67–81.

LIANG, K.-Y. AND ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

LUNETTA, K. L., FARAONE, S. V., BIEDERMAN, J. AND LAIRD, N. M. (2000). Family-based tests of association and linkage that use unaffected sibs, covariates, and interactions. *American Journal of Human Genetics* **66**, 605–614.

MARTIN, E. R., MONKS, S. A., WARREN, L. L. AND KAPLAN, N. L. (2000). A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *American Journal of Human Genetics* **67**, 146–154.

MARTINEZ, F. D., GRAVES, P. E., BALDINI, M., SOLOMON, S. AND ERICKSON, R. (1997). Association between genetic polymorphisms of the beta-2 adrenoceptor and response to albuterol in children with and without a history of wheezing. *Journal of Clinical Investigation* **100**, 3184–3188.

RABINOWITZ, D. (1997). A transmission disequilibrium test for quantitative trait loci. *Human Heredity* **47**, 1113–1118.

RABINOWITZ, D. AND LAIRD, N. M. (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and‘arbitrary missing marker information. *Human Heredity* **50**, 211–223.

SCHAID, D. J. AND SOMMER, S. S. (1994). Comparison of statistics for candidate-gene association studies using cases and parents. *American Journal of Human Genetics* **55**, 402–409.

SCHAID, D. J. AND SOMMER, S. S. (1996). General score tests for association of genetic markers with disease using cases and their parents. *Genetic Epidemiology* **13**, 423–449.

SPIELMAN, R. S., MCGINNIS, R. E. AND EWENS, W. J. (1993). Transmisson test for linkage disequilibrium: the insulin gene region and insulin-dependent Diabetes Mellitus (iddm). *American Journal of Human Genetics* **52**, 506–516.

SUMMERHILL, E., LEAVITT, S. A., GIDLEY, H., PARRY, R., SOLWAY, J. AND OBER, C. (2000). beta(2)-adrenergic receptor arg16/arg16 genotype is associated with reduced lung function, but not with asthma, in the hutterites. *American Journal of Respiratory and Critical Care Medicine* **162**, 599–602.

THOMSON, G. (1995). Analysis of complex human genetic traits: an ordered-notation method and new tests for mode of inheritance. *American Journal of Human Genetics* **57**, 474–486.

TURKI, J., PAK, J., GREEN, S. A., MARTIN, R. J. AND LIGGETT, S. B. (1995). Genetic polymorphisms of the beta 2-adrenergic receptor in nocturnal and nonnocturnal asthma. evidence that gly16 correlates with the nocturnal phenotype. *Journal of Clinical Investigation* **95**, 1635–1641.

WHITTAKER, J. C. AND LEWIS, C. M. (1998). Power comparisons of the transmission/disequilibrium test and sib-transmission/disequilibrium-test statisitics. *American Journal of Human Genetics* **65**, 578–580.

ZHAO, H. (2000). Family-based association studies. *Statistical Methods in Medical Research* **9**, 563–587.

[*Received October 31, 2001; revised March 11, 2002; accepted for publication May 13, 2002*]