

# *Statistical Applications in Genetics and Molecular Biology*

---

*Volume 8, Issue 1*

2009

*Article 17*

---

## Impact of Population Stratification on Family-Based Association Tests with Longitudinal Measurements

Xiao Ding\*      Scott Weiss†      Benjamin Raby‡  
Christoph Lange\*\*      Nan M. Laird††

\*Harvard School of Public Health, xiao.ding@post.harvard.edu

†Harvard Medical School, scott.weiss@channing.harvard.edu

‡Harvard Medical School, benjamin.raby@channing.harvard.edu

\*\*Harvard School of Public Health, clange@hsph.harvard.edu

††Harvard School of Public Health, laird@hsph.harvard.edu

Copyright ©2009 The Berkeley Electronic Press. All rights reserved.

# Impact of Population Stratification on Family-Based Association Tests with Longitudinal Measurements\*

Xiao Ding, Scott Weiss, Benjamin Raby, Christoph Lange, and Nan M. Laird

## Abstract

Several family-based approaches for testing genetic association with traits obtained from longitudinal or repeated measurement studies have been previously proposed. These approaches utilize the multivariate data more efficiently by using estimated optimal weights to combine univariate tests. We show that these FBAT approaches are still robust against hidden population stratification, but their power can be heavily affected since the estimated weights might provide poor approximation of the true theoretical optimal weights with the presence of population stratification. We introduce a permutation-based approach FBAT-MinP and an equal combination approach FBAT-EW, both of which do not involve the use of estimated weights. Through simulation studies, FBAT-MinP and FBAT-EW are shown to be powerful even in the presence of population stratification, when other approaches may substantially lose their power. An application of these approaches to the Childhood Asthma Management Program (CAMP) study data for testing an association between body mass index and a previously reported candidate SNP is given as an example.

**KEYWORDS:** family-based association test, population stratification, longitudinal phenotypes

---

\*We thank all subjects for their ongoing participation in this study. We acknowledge the CAMP investigators and research team, supported by the National Heart, Lung, and Blood Institute (NHLBI), for collection of CAMP Genetic Ancillary Study data. The CAMP Genetics Ancillary Study is supported by the NHLBI, N01-HR-16049. Additional support for this research came from grants P50 HL67664 and T32 HL07427 from the National Institutes of Health and the NHLBI. All work on data from the CAMP Genetics Ancillary Study was conducted at the Channing Laboratory and the Brigham and Women's Hospital under appropriate CAMP policies and human subject's protections. This study was supported by the National Institutes of Health (NIH) grants GM 029745 and MH 05932.

# 1 Introduction

For a genetic association study, population stratification is present when the population is comprised of several subpopulations, and the allele frequency of interest differs in each subpopulation (Ewens and Spielam, 1995; Risch, 2000; Lander and Schork, 1994; Pritchard and Donnelly, 2001). Although most population stratification occurs when there are multiple races or ethnicities in the study population (Risch et al., 2002), significant population stratification was reported even within an apparently homogeneous North American population of European ancestry (Campbell et al., 2005).

Population stratification is mainly thought to be a concern for case-control studies. Since family-based association studies are robust against population stratification, they are often used to replicate associations found with case-control studies. However family-based tests can lose power in the presence of population stratification (Lewinger and Bull, 2006; He et al., 2008), and it is therefore particularly important to address the impact of population stratification on family-based studies in order to ensure these studies have sufficient power to detect the target genetic association.

For studies with phenotypes measured longitudinally or repeatedly, several family-based association test (FBAT) approaches have been introduced which use population information to enhance the power (Lange et al., 2004; Ding et al., 2009). All these approaches combine univariate tests to construct more powerful global test, by using different estimated weights to approximate the theoretical optimal weights. We evaluate the impact of population stratification on the power of these FBAT approaches and find that these estimated weights may no longer provide an appropriate approximation of the optimal weights in the presence of population stratification. We suggest that equal weights can be used to combine univariate tests to provide an alternative testing approach, which can be robust against population stratification. In addition, by extending the idea of combining tests of multiple markers, we propose a permutation-based approach FBAT-Min P, which is expected to be a more powerful approach than Bonferroni correction.

# 2 Methods

Suppose there are  $N$  families. For simplicity, assume we have parents with one offspring (trios); the results can be easily generalized to other family structures (Rabinowitz and Laird, 2000). We denote the vector containing all

$m$  phenotypic observations for each offspring by  $\tilde{Y}_i = (Y_{i1}, \dots, Y_{im})^T$ , where  $Y_{ij}$  is the  $j$ -th phenotype for the  $i$ -th offspring. The standard biometric model (Falconer, 1997) describing a single phenotype as a function of the genotype is

$$E(Y_{ij}|X_i = x_i) = \mu_j + \alpha_j \times x_i, \quad (1)$$

where  $\alpha_j$  is the genetic effect for the  $j$ -th measurement, and  $X_i$  denotes the coding of the marker genotype of the  $i$ -th offspring. The vector containing all traits for each offspring can be expressed as  $\tilde{T}_i = (T_{i1}, \dots, T_{im})^T$ , where  $T_{ij}$  is the  $j$ -th trait for the  $i$ -th offspring. Here  $T_{ij}$  is a function of the phenotype  $Y_{ij}$  for example,  $T_{ij} = Y_{ij} - \bar{Y}_{.j}$  or  $Y_{ij}$  adjusted for covariates (Lunetta, 2000).

For the  $j$ -th measurement, the univariate family-based association test (FBAT) statistic (Rabinowitz and Laird, 2000) can be written as

$$S_j = \sum_{i=1}^N T_{ij}[X_i - E(X_i|P_i)], \quad (2)$$

where  $E(X_i|P_i)$  and  $Var(X_i|P_i)$  denote the expectation and variance of the marker score computed under the null hypothesis (no genetic association), conditional on the parental genotypes  $P_i$ . With large samples, the vector containing all univariate test statistics  $\tilde{S} = (S_1, S_2, \dots, S_m)^T$  asymptotically follows a multivariate normal distribution  $N(\tilde{0}_m, \Sigma_0)$  under  $H_0$  (Lange et al., 2003b). Here  $\tilde{0}_m$  is an  $m$ -dimensional vector of zeroes and  $\Sigma_0$  is the variance-covariance matrix of those univariate test statistics:

$$\Sigma_0 = Var(\tilde{S}|H_0) = \sum_{i=1}^N \tilde{T}_i \tilde{T}_i^T Var(X_i|P_i). \quad (3)$$

Several approaches have been introduced to utilize the multivariate data efficiently to test for genetic association in family-based studies. Lange et al. (2004) developed the FBAT-PC approach, which is an extension of FBAT for traits that are measured longitudinally or repeatedly over time. Based on generalized principle component analysis, FBAT-PC uses a weighted linear combination of the measurements to construct an overall phenotype with maximal locus-specific heritability. To avoid biasing the significance level of any subsequent tests, Lange and Laird (2003), Lange et al. (2003a) proposed the Conditional Mean Model (CMM) to estimate the optimal weights to use in FBAT-PC. In equation (1), we replace the observed marker score  $x_i$  by the expected marker score  $E(X_i|P_i)$ ,

$$E(Y_{ij}) = \mu_j + \alpha_j \times E(X_i|P_i). \quad (4)$$

Ding et al. (2009) introduced FBAT-PCM as a modification to FBAT-PC with higher power, along with two other approaches: FBAT-LC and FBAT-LCC, which have more power in some circumstances.

Furthermore, all three of these statistics can be expressed as a weighted combination of those univariate tests  $S_j$ , with different methods to compute the weights:

$$Z_{FBAT-LC} = \frac{\tilde{q}^T \tilde{S}}{\sqrt{\tilde{q}^T \Sigma_0 \tilde{q}}}, \quad (5)$$

$$Z_{FBAT-LCC} = \frac{(\Sigma_0^{-1} \tilde{q})^T \tilde{S}}{\sqrt{(\Sigma_0^{-1} \tilde{q})^T \Sigma_0 (\Sigma_0^{-1} \tilde{q})}}, \quad (6)$$

$$Z_{FBAT-PCM} = \frac{(\hat{V}_P^{-1} \hat{\alpha})^T \tilde{S}}{\sqrt{(\hat{V}_P^{-1} \hat{\alpha})^T \Sigma_0 (\hat{V}_P^{-1} \hat{\alpha})}}, \quad (7)$$

where  $\tilde{q} = (\frac{\hat{\alpha}}{SE(\hat{\alpha})})$ , and  $V_P = Var(\tilde{Y}_i | X_i = x_i)$  is the phenotypic residual variance-covariance matrix. Note that  $\hat{\alpha}$  is estimated by the conditional mean model (equation 4), and all the weights used in equation (5) to (7) are different approximation of the theoretical optimal weights (Ding et al., 2009). Thus it is important that  $\hat{\alpha}$  is a good estimate of the true genetic effect as in equation (1).

### Use equal weights to combine univariate statistics: FBAT-EW

The weight vectors for FBAT-LC, FBAT-LCC and FBAT-PCM are all estimated by the conditional mean model. Although the CMM will not bias any subsequent tests, it can introduce noise, especially when there is hidden population substructure. Therefore one intuitive question may arise: do we really gain by estimating those weights via the CMM; what if we simply use some constants as our weights?

Without any previous information, the natural way of choosing the constants is to let the weights be all equal. In other words, instead of using estimates like  $\tilde{q}$ ,  $\Sigma_0^{-1} \tilde{q}$  and  $\hat{V}_P^{-1} \hat{\alpha}$  in equations (5)-(7), we just assign the weight vector to be  $\tilde{1}_m$  (a vector of ones) and name the statistic as FBAT-EW,

$$Z_{FBAT-EW} = \frac{\tilde{1}_m^T \tilde{S}}{\sqrt{\tilde{1}_m^T \Sigma_0 \tilde{1}_m}}. \quad (8)$$

Note that this is the same as taking the summation of each individual test  $S_j$  as the test statistic and normalizing by its standard error. Under the alternative, those weights for FBAT-LC, FBAT-LCC and FBAT-PCM should be positively correlated with  $\tilde{S}$  (Ding et al., 2009), intuitively because  $\tilde{\alpha}$  is positively correlated with  $\tilde{S}$ . Here since the weights are chosen to be vector of ones, we can no longer guarantee a positive correlation between the weights and the univariate tests, thus FBAT-EW is a two-sided test.

Assuming that all the underlying genetic effects are equal, i.e.,  $\alpha_1 = \alpha_2 = \dots = \alpha_m$  and the phenotypic residual variance matrix  $V_P$  is compound symmetry, the optimal weights to combine univariate tests can be easily shown to be approximately proportional to  $\tilde{1}_m$ . Therefore, when the true genetic effects for different measurement points are nearly equal, and the variance-covariance matrix is close to compound symmetry, FBAT-EW is expected to work well. When using equal weights, we can avoid the possible non-robustness of estimating the weights via the conditional mean model. But at the same time, we make the assumption of equal genetic effects and a compound symmetry variance-covariance matrix, which may or may not be appropriate, depending upon the underlying model. Furthermore, FBAT-EW may not be appropriate when the correlation among traits is negative.

### Permutation test based on univariate statistics: FBAT-Min P

For the univariate test statistic  $S_j$  in equation (1), we can standardize it as

$$Z_j = \frac{S_j}{\sqrt{V_j}} = \frac{\sum_{i=1}^N T_{ij} [X_i - E(X_i|P_i)]}{\sqrt{\sum_{i=1}^N T_{ij}^2 [Var(X_i|P_i)]}}. \quad (9)$$

Letting  $C_0$  be the correlation matrix corresponding to the variance-covariance matrix  $\Sigma_0$ , it is easy to show that  $\tilde{Z} = (Z_1, Z_2, \dots, Z_m)^T$  asymptotically follows multivariate normal distribution,  $N(\tilde{0}_m, C_0)$ , under the null hypothesis.

In order to get a one degree of freedom test and gain more power than Bonferroni correction, we define our statistic as the most significant univariate test (whose p-value is the minimum):

$$Z_{FBAT-MinP} = \max(|Z_1|, |Z_2|, \dots, |Z_m|). \quad (10)$$

Though the distribution of  $Z_{FBAT-MinP}$  under the null hypothesis is theoretically difficult to derive, the corresponding p-value can be easily obtained via Monte Carlo permutation. First, we generate  $K$  (a sufficiently

large number) random  $m$ -dimensional vectors  $\tilde{U}_1, \dots, \tilde{U}_K$  that follow a multivariate normal distribution  $N(\tilde{0}_m, C_0)$ . Second, for each vector  $\tilde{U}_k$ ,  $k = 1, \dots, K$ , we find the maximum absolute value of its  $m$  elements and record it as  $U_{max}^k = \max_{1 \leq l \leq m} |U_{kl}|$ . Therefore, the set  $F = \{U_{max}^1, U_{max}^2, \dots, U_{max}^K\}$  is a random sample from the distribution of the standardized test statistic with the minimum univariate  $p$ -value. Last, we determine how many  $U_{max}^k$  are actually bigger or equal to  $Z_{FBAT-MinP}$ , and then divide the number by  $K$  to get the empirical  $p$ -value:

$$P_{FBAT-MinP} = \sum_{k=1}^K I(U_{max}^k \geq Z_{FBAT-MinP}) / K. \quad (11)$$

### 3 Simulation

In our simulations, the marker of interest is a bi-allelic locus. Assuming an additive genetic model, the parental genotypes P1 and P2 are generated by drawing from a binomial distribution  $B(2, p)$  where  $p$  is the minor allele frequency (MAF) of the target allele in the population. The genotype  $X$  of the offspring is obtained by simulated Mendelian transmission based on the parental genotypes P1 and P2. For each offspring, the same type of phenotype is measured 6 times. The 6-dimensional phenotypic vector is a random sample from a multivariate normal distribution

$$\tilde{Y}_i = (y_{i1}, \dots, y_{i6})^T \sim MVN(\tilde{\mu} + (\alpha_1, \dots, \alpha_6)^T X_i, V_P), \quad (12)$$

where  $V_P$  is the phenotypic variance-covariance matrix,  $\tilde{\mu} = 15 \times \tilde{1}_6$  is the phenotypic mean and  $\alpha_1, \dots, \alpha_6$  are the genetic effects for measurement 1 to 6, respectively.

The simulation is repeated 10,000 times, in each replicate, 400 trios are generated for analysis. The power of each approach is estimated by the proportion of the number of times when the test statistic is significant at  $\alpha$  level=0.05 (alpha level=0.01 and 0.001 are also studied, and the results are very similar). We only report results for MAF  $p=0.2$ , as results for other values are very similar. Since the power of a statistical test heavily depends upon the true underlying model, we perform our simulations under several different models for the genetic effects  $\alpha_1, \dots, \alpha_6$ . In all the models, the variances at each measurement are set to  $\sigma_i^2 = 1, i = 1, \dots, 6$ , while the correlation matrix  $C_P$  is compound symmetry with various correlation values.

In other words,  $C_p = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & & & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}$ , where  $\rho$  is the correlation among different measurements for the same subject. Therefore, we have

$$V_p = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_6 \end{pmatrix} \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & & & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_6 \end{pmatrix}.$$

Other correlation structures taken from some actual data set have also been used but are not reported in this paper since the results are quite similar to the compound symmetry model.

**Model 1: no genetic effect at any measurement point**

Under the null hypothesis, there is no genetic association at all (i.e. the genetic effect is zero for any of the six measurement points), so the phenotypes are generated from  $\alpha_i = 0, i = 1, \dots, 6$ .

**Model 2: same genetic effects across all measurement points**

In this model, we assume that  $\alpha_i = \alpha_h, i = 1, \dots, 6$ , where  $\alpha_h$  is the genetic effect size that corresponds to the heritability  $h^2$  (Falconer, 1997), i.e.,  $\alpha_h = \sqrt{\frac{h^2}{2p(1-p)(1-h^2)}}$  for an additive genetic model. In model 2 and model 3,  $h^2$  is always set to be 0.01.

**Model 3: arbitrary effects for different measurement points**

Here the values of  $\alpha_1, \dots, \alpha_6$  are given by

$$\alpha_j \sim U(0, 2\alpha_h), \tag{13}$$

where  $U$  is the uniform distribution on the interval. Since the mean of the uniform distribution is  $\alpha_h$ , the average genetic effect here is also  $\alpha_h$ , with the average univariate heritability equal to 0.01.

**Consideration of population stratification**

In order to study the influence of population stratification on the power of these approaches, we modify our simulations to include possible population substructure. Instead of choosing all 400 trios from the same population of MAF  $p=0.2$  and  $\tilde{\mu} = 15 \times \tilde{1}_{15}$  in equation (12), now we select 200 trios from subpopulation 1 and 200 trios from subpopulation 2 ( $p_1 = 0.1, \tilde{\mu}_1 = (15 + \beta_1) \times \tilde{1}_{15}$  in subpopulation 1 and  $p_2 = 0.3, \tilde{\mu}_2 = (15 + \beta_2) \times \tilde{1}_{15}$  in subpopulation 2).



For model 1, 2 and 3, we consider two extreme cases of possible population substructure:  $\beta_1 = -\alpha_h, \beta_2 = +\alpha_h$  for case 1, and  $\beta_1 = +\alpha_h, \beta_2 = -\alpha_h$  for case 2. For the first case, the correlation between phenotypic means and MAF of the two subpopulations is positive, which means the false positive results caused by population stratification are in the same direction as the true genetic effects. On the contrary, in case 2, the correlation is negative and the false signals caused by population stratification tend to cancel out the true genetic effects.

Table 1: Type-I error rates of FBAT-EW and FBAT-Min P under the null

**without any population stratification**

<b>Correlation</b>	0	0.1	0.6	0.9
<b>FBAT-EW</b>	0.050	0.050	0.052	0.049
<b>FBAT-Min P</b>	0.052	0.049	0.048	0.049

**positive population stratification (case 1)**

<b>Correlation</b>	0	0.1	0.6	0.9
<b>FBAT-PCM</b>	0.048	0.047	0.054	0.050
<b>FBAT-LC</b>	0.049	0.047	0.052	0.052
<b>FBAT-LCC</b>	0.048	0.049	0.052	0.050
<b>FBAT-EW</b>	0.049	0.049	0.053	0.051
<b>FBAT-Min P</b>	0.050	0.043	0.048	0.052

**negative population stratification (case 2)**

<b>Correlation</b>	0	0.1	0.6	0.9
<b>FBAT-PCM</b>	0.048	0.050	0.051	0.046
<b>FBAT-LC</b>	0.049	0.046	0.050	0.043
<b>FBAT-LCC</b>	0.049	0.048	0.051	0.047
<b>FBAT-EW</b>	0.050	0.046	0.051	0.052
<b>FBAT-Min P</b>	0.048	0.047	0.054	0.052

## 4 Results

Regardless of population stratification, the type-I error rates of FBAT-PCM, FBAT-LC and FBAT-LCC are all well maintained (Ding et al., 2009). For various values of the correlation  $\rho$ , the type-I error rates of FBAT-EW and FBAT-Min P under the null (model 1) are also well maintained (Table 1).

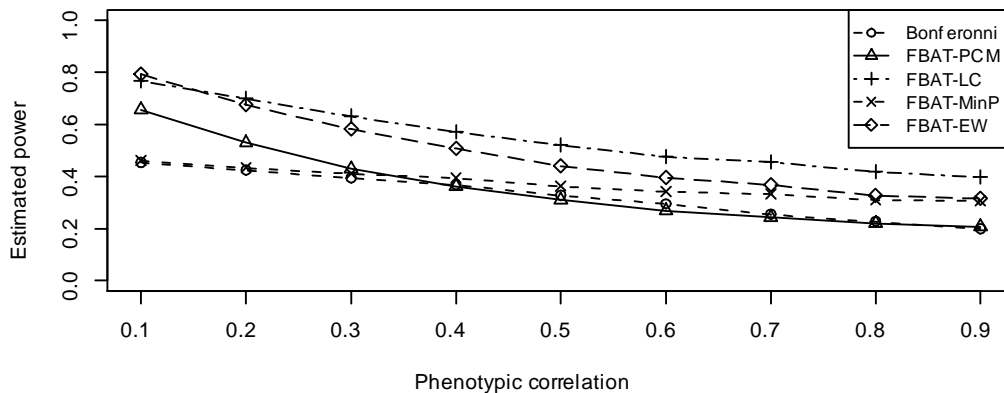
When there is no population stratification, Figure 1 and 4 show the estimated power of FBAT approaches, under models 2 and 3, respectively. When the variances are same, FBAT-LCC and FBAT-PCM are very similar to each other (Ding et al., 2009), so for simplicity we do not show FBAT-LCC. From these two figures, we see that FBAT-Min P always has higher power than Bonferroni correction, under either model 2 or model 3. Furthermore, since FBAT-Min P takes the correlation among traits into account, its power gain over Bonferroni correction increases when the correlation  $\rho$  increases. As for FBAT-EW, when the genetic effect sizes are all equal (model 2), it is more powerful than FBAT-PCM, FBAT-Min P and Bonferroni correction. FBAT-EW performs only slightly worse than FBAT-LC, which is basically due to the fact that FBAT-EW is a two-sided test, while FBAT-LC is one-sided. On the other hand, when the genetic effect sizes differ substantially (model 3), FBAT-EW is no longer a very powerful approach, especially when the correlation among traits is high.

We find that population substructure can affect the power of these approaches in either direction. In other words, population substructure can either increase (positive population stratification, i.e., case 1) or decrease (case 2 for negative stratification) the power, depending upon whether the false positive results caused by population substructure are in the same direction as the true genetic effects and which FBAT approach is used. In case 1, the hidden population substructure tends to increase the power of those weighted FBAT approaches, but not some other FBAT approaches such as FBAT-EW and FBAT-MinP. In case 2, the hidden population substructure is expected to reduce the power of all FBAT approaches. In addition, case 2 is the worst situation whereby population substructure can weaken the signal of the true genetic association. Therefore, we focus on case 2 to compare the power of these FBAT approaches under the worst case of hidden population substructure.

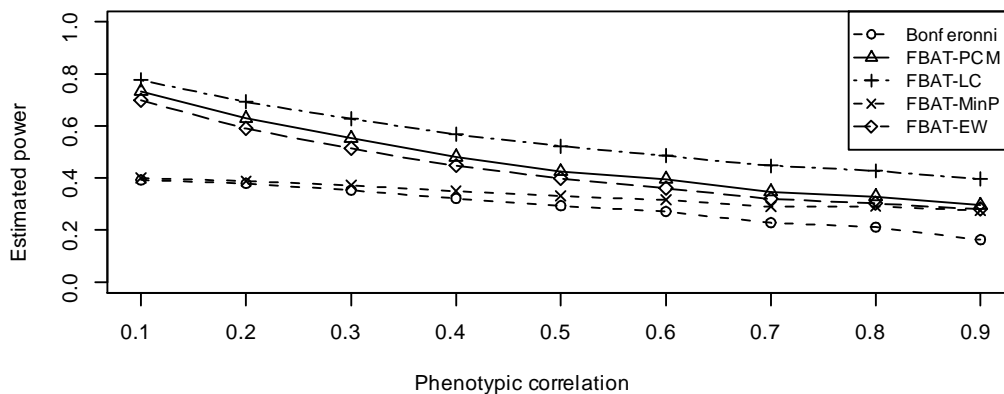
Figure 2 and 5 show the estimated power of these approaches under model 2 and model 3, respectively, when positive population stratification (case 1) exists.

By comparing Figure 2 to Figure 1, as well as Figure 5 to Figure 4, we find that the positive population stratification can slightly improve the

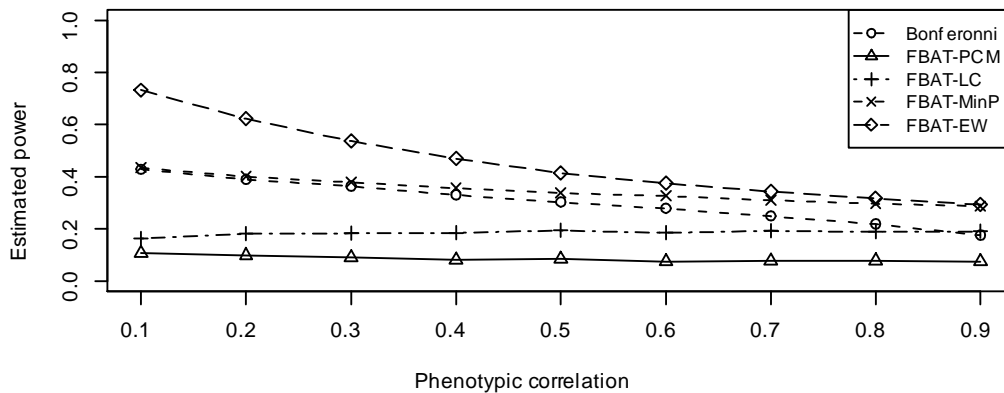
**Figure 1: Estimated power when genetic effect sizes are same (model 2) and there is no population stratification**



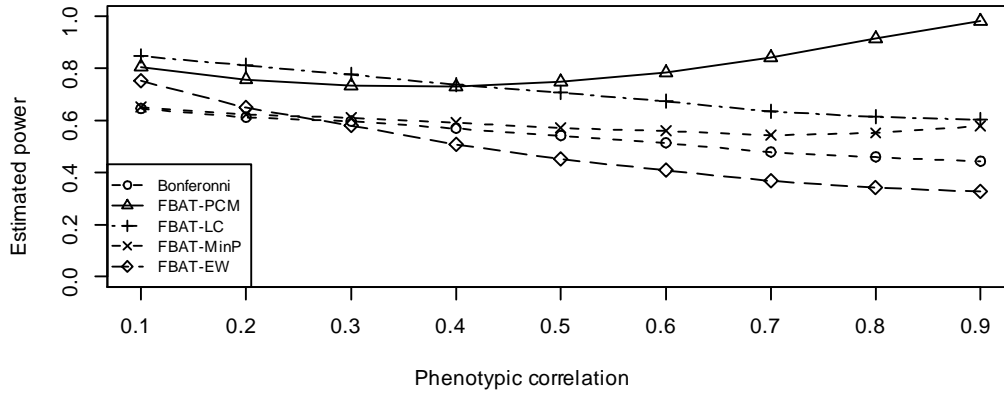
**Figure 2: Estimated power when genetic effect sizes are same (model 2), under positive population stratification**



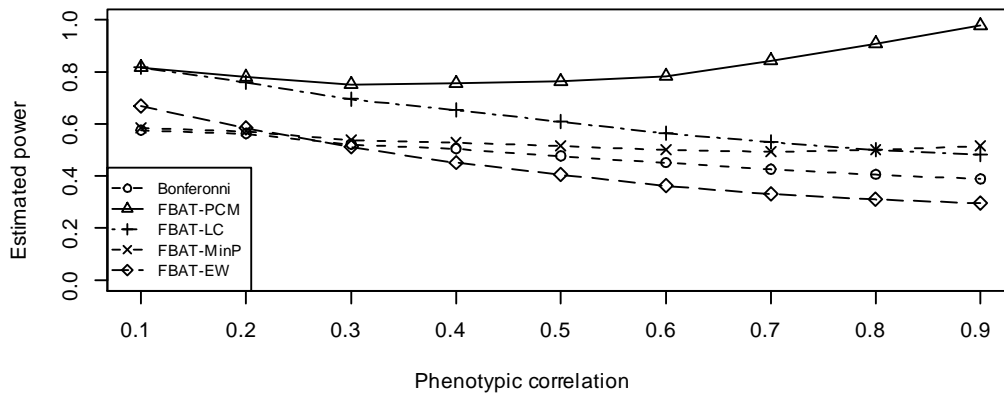
**Figure 3: Estimated power when genetic effect sizes are same (model 2), under negative population stratification**



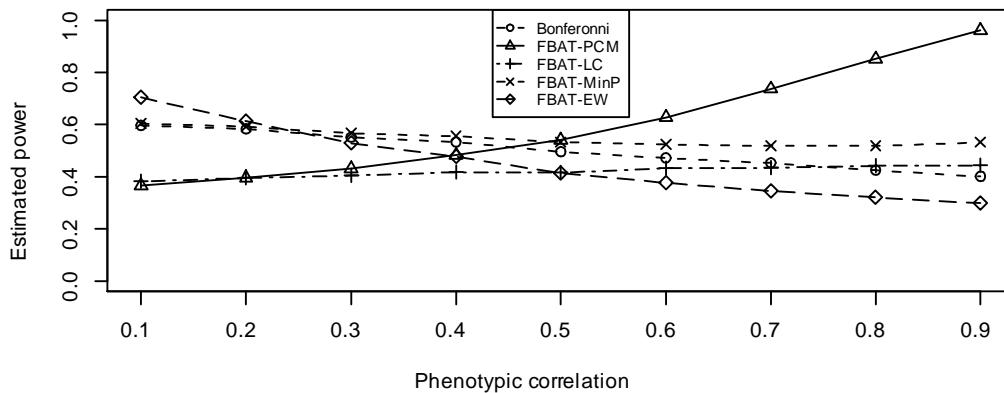
**Figure 4: Estimated power when genetic effects uniformly distribute (model 3) and there is no population stratification**



**Figure 5: Estimated power when genetic effects uniformly distribute (model 3), under positive population stratification**



**Figure 6: Estimated power when genetic effects uniformly distribute (model 3), under negative population stratification**



power of FBAT-PCM and FBAT-LC, especially under model 2. Since the false positive results caused by population substructure are in the same direction as true genetic effect, they tends to enhance the overall signal detected by the weighted FBAT approaches, i.e., FBAT-PCM and FBAT-LC. On the other hand, positive population stratification slightly reduce the power of FBAT-EW and FBAT-MinP. This is because those univariate FBAT statistics may slightly lose power due to positive population stratification.

Figure 3 and 6 show the estimated power of these approaches under model 2 and model 3, respectively, when negative population stratification (case 2) exists.

By comparing Figure 3 to Figure 1, as well as Figure 6 to Figure 4, we find that the impact on the power of FBAT-PCM and FBAT-LC is substantial. This is due to the fact that both FBAT-PCM and FBAT-LC use the conditional mean model, which might provide poor approximation (or even approximation in the wrong direction) of the true underlying genetic effects with the presence of population stratification. An exception is that FBAT-PCM consistently does well for very high values of  $\rho$  when the genetic effect sizes are different (model 3). As expected, the impact of population stratification on FBAT-Min P and FBAT-EW is very limited, since neither approach involves the use of the conditional mean model. FBAT-Min P remains a powerful test with the presence of population stratification, and still holds a noticeable gain of power compared to Bonferroni correction. FBAT-EW has the highest power among all the approaches when the genetic effect sizes are all equal, or the correlation  $\rho$  is very low.

Table 2: Summary of the power of FBAT-EW and FBAT-Min P

	without any population stratification	with positive population stratification	with negative population stratification
<b>FBAT EW</b>	has good power when $\alpha_j$ are all equal; but not very powerful when $\alpha_j$ are different unless the correlation is low	its power is only slightly affected; has good power when $\alpha_j$ are all equal, but outperformed by FBAT-PCM and FBAT-LC	its power is only slightly affected; the most powerful approach when either $\alpha_j$ are all equal or $\rho$ is low
<b>FBAT Min P</b>	generally a powerful approach, especially when effects $\alpha_j$ vary; has noticeable power gain over Bonferroni correction, especially when $\rho$ is high	its power is only slightly affected; outperformed by FBAT-PCM and FBAT-LC	its power is only slightly affected; remain powerful under all situations

When the conditional mean model fails to approximate the true genetic model well, the FBAT approaches which use the conditional mean model may lose power since their weights are no longer good estimates of the unknown optimal weights. Theoretically, it is possible to adjust the conditional mean model for population substructure with approaches such as Eigenstrat (Price et al., 2006), suggesting that all the FBAT approaches involving the conditional mean model can maintain their power even when population substructure exists. However, in evaluating the impact of population substructure, remember that we considered a 'worst possible case', where the correlation between allele frequency and mean phenotype worked to negate the true phenotype-genotype association in the population. A correlation in the opposite direction would increase power. In practice, it will be difficult to predict the sign and extent of any correlation between marker allele frequencies and traits.

## 5 Data analysis

We test for the association between SNP rs7566605 and 15 longitudinal measures of BMI over 48 months for 212 parent-child trios from the Childhood Asthma Management Program (CAMP) study. SNP rs7566605 is located on chromosome 2q14.2 near the INSIG2 gene and has been reported to be associated with obesity in several populations (Herbert et al., 2006; Lyon et al., 2007).

Fulker et al. (1999) suggested that the genetic effect of a quantitative trait could be decomposed into between-family (b) and within-family (w) components. For family-based trio data, the model can be written as

$$E(Y_{ij}) = \mu_j + \alpha_{bj} \times E(X_i|P_i) + \alpha_{wj} \times [X_i - E(X_i|P_i)], \quad (14)$$

where  $\alpha_{bj}$  and  $\alpha_{wj}$  are the between-family and within-family genetic effect sizes, respectively. Note that if there is no population stratification,  $\hat{\alpha}_{bj}$  and  $\hat{\alpha}_{wj}$  are both expected to be unbiased estimates of the true underline genetic effect, therefore the model reduces to equation (1). Furthermore, fitting the conditional mean model should give estimates nearly equivalent to  $\hat{\alpha}_{bj}$ , since  $E(X)$  and  $X - E(X)$  are orthogonal in expectation (i.e.  $E\{Cov(E(X|P), X - E(X|P))\} = 0$ ) [Falconer 1997], assuming an unselected sample). However, when sampling from a stratified population, even though  $\hat{\alpha}_{wj}$  will remain unbiased,  $\hat{\alpha}_{bj}$  may be biased. As shown in Table 3, between-family estimates  $\hat{\alpha}_{bj}$  and within-family estimates  $\hat{\alpha}_{wj}$  not only differ greatly in the values, but also hold opposite signs. In other words, the conditional

mean model is no longer a reasonable approximate of the true genetic effects.

Table 3: Estimators of genetic effect at each time point, based on Fulker model

Model	Between-family components		Within-family components	
	$\hat{\alpha}_b$	$s.e.(\hat{\alpha}_b)$	$\hat{\alpha}_w$	$s.e.(\hat{\alpha}_w)$
<b>BMI-PRE</b>	-0.63	0.94	3.15	0.92
<b>BMI-RZ</b>	-0.43	0.96	2.98	0.94
<b>BMI-M2</b>	-0.40	0.98	3.15	0.95
<b>BMI-M4</b>	-0.43	0.99	3.08	0.96
<b>BMI-M8</b>	-0.45	1.04	2.88	1.02
<b>BMI-M12</b>	-0.47	1.06	2.89	1.03
<b>BMI-M16</b>	-0.14	1.08	2.81	1.05
<b>BMI-M20</b>	-0.25	1.12	3.01	1.09
<b>BMI-M24</b>	-0.18	1.15	3.09	1.12
<b>BMI-M28</b>	-0.17	1.15	2.80	1.12
<b>BMI-M32</b>	-0.11	1.19	2.90	1.16
<b>BMI-M36</b>	-0.19	1.20	2.76	1.17
<b>BMI-M40</b>	-0.29	1.25	2.69	1.21
<b>BMI-M44</b>	-0.27	1.30	2.47	1.27
<b>BMI-M48</b>	-0.76	1.30	2.39	1.26

Here BMI-PRE is BMI measured at two months before the randomization, BMI-RZ means BMI measured at randomization, and BMI-M2, M4,  $\dots$  represent BMI measured at 2, 4,  $\dots$  months of follow up.

For these data, FBAT-PC, FBAT-PCM, FBAT-LC and FBAT-LCC all fail to detect the overall genetic association, while FBAT-Min P and FBAT-EW give significant results. The computation of both FBAT-Min P and FBAT-EW does not involve the conditional mean model at all, while all the other approaches use it. Therefore, when hidden population substructure is a concern, weighted FBAT approaches such as FBAT-PCM, FBAT-LC and FBAT-LCC may lose their power to detect an existing association, while our new approach FBAT-EW and FBAT-Min P give significant results.

## 6 Discussion

In this paper, we evaluate the impact of population stratification on the family-based association studies. Because of the robustness of univariate FBAT statistic against population stratification, the false positive rates are

Table 4: Testing for association between rs7566605 and BMI in CAMP data

	P-value
<b>FBAT-PCM</b>	0.58
<b>FBAT-LC</b>	0.99
<b>FBAT-LCC</b>	0.73
<b>Bonferroni correction</b>	0.14
<b>FBAT-EW</b>	0.028
<b>FBAT-Min P</b>	0.022

always well maintained for the longitudinal setting. But the power of FBAT approaches to detect true genetic association can be either increased or decreased by the existing population stratification in the study samples, depending upon the direction of the population stratification bias. For several FBAT approaches previously introduced to enhance the power of family-based studies with longitudinal phenotypes, we show that they can lose their power substantially when negative population stratification tends to cancel out the true signal of genetic effects. Here we introduce a permutation-based test FBAT-Min P and an equally weighted combination test FBAT-EW, both of which remain to be powerful approaches when population stratification is present.

As shown in the simulations and the analysis of the CAMP data, both FBAT-EW and FBAT-Min P do not need to fit the conditional mean model and are robust against hidden population substructure, so they are very good approaches to start with, regardless of whether there is a concern of population stratification. Furthermore, if we expect the genetic effects for different measurement points to be about the same, FBAT-EW is always a simple and powerful test. When the genetic effect size differs substantially, FBAT-PCM is very powerful when the correlation among traits is high, regardless of the presence or absence of hidden population stratification under our setting.

The computation of all these FBAT approaches is straightforward once you have all the univariate FBAT test statistics. In addition, univariate FBAT, FBAT-LC and FBAT-Min P have been implemented in the software package FBAT and is freely available at <http://www.biostat.harvard.edu/~fbat/default.html>; FBAT-PC and FBAT-PCM have been implemented in the software package PBAT and is freely available at <http://www.biostat.harvard.edu/~clange/default.html>.



## References

- The Childhood Asthma Management Program Research Group. (1999) The Childhood Asthma Management Program (CAMP): design, rationale, and methods. *Controlled Clinical Trials* 20(1), 91-120.
- The Childhood Asthma Management Program Research Group. (2000) Long-term effects of budesonide or nedocromil in children with asthma. *The New England journal of medicine* 343(15), 1054-63.
- Campbell, C., Ogburn, E., Lunetta, K., Lyon, H., Freedman, M., Groop, L., Altshuler, D., Ardlie, K., and Hirschhorn, J. (2005). Demonstrating stratification in a European American population. *Natural Genetics* 37(8), 868-872.
- Ding, X., Lange, C., Xu, X. and Laird, N. (2009) New powerful approaches for family-based association tests with longitudinal measurements. *Annals of Human Genetics* 73, 74-83.
- Ewens, W. and Spielman, S. (1995) The Transmission/Disequilibrium Test: History, Subdivision and Admixture. *American Journal of Human Genetics* 57, 455-464.
- Falconer, D. and Mackay, T. (1997) Introduction to quantitative genetics. *Longman*.
- Fulker, D., Cherny, S., Sham, P., and Hewitt, J. (1999) Combined linkage and association analysis for quantitative traits. *American Journal of Human Genetics* 64, 259-267.
- He, Y., Jiang, R., Bergen, A., Swan, G. and Jin, L. (2008) Correlation of Population Parameters leading to Power Differences in Association Studies with Population Stratification. *Annals of Human Genetics* 72, 801-811.
- Herbert, A., Gerry, N., McQueen, M., Heid, I., Pfeufer, A., Illig, T., Wichmann, H., Meitinger, T., Hunter, D., Hu, F., Colditz, G., Hinney, A., Hebebrand, J., Koberwitz, K., Zhu, X., Cooper, R., Ardlie, K., Lyon, H., Hirschhorn, J., Laird, N., Lenburg, M., Lange, C., and Christman M. (2006) A common genetic variant is associated with adult and childhood obesity. *Science* 312, 279-283.
- Laird, N. and Lange, C. (2006) Family-Based designs in the age of large-scale gene-association studies. *Nature Reviews/Genetics* 7, 385-394.
- Lander, E. and Schork, N. (1994) Genetic dissection of complex traits. *Science* 265, 2037-2048.

- Lange, C and Laird, N. (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., Lyon, H., DeMeo, D., Raby, B., Silverman, A., and Weiss, S. (2003) A new powerful non-parametric two-stage approach for testing multiple phenotypes in family-based association studies. *Human Heredity* 56, 10-17.
- Lange, C., Silverman, E., Xu, X., Weiss, S. and Laird, N. (2003) A multivariate family-based association test using generalized estimating equations: FBAT-GEE. *Biostatistics* 4, 195-206.
- Lange, C., and Laird, N. (2003) Analytical sample size and power calculations for a general class of family-based association tests: Dichotomous traits. *American Journal of Human Genetics* 23, 165-180.
- Lange, C., Andrew, T., MacGregor, A., Lyon, H., Raby, B., DeMeo, D., Murphy, A., Silverman, A., Weiss, S., and Laird, N. (2004) A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Statistical Applications in Genetics and Molecular Biology* 3, Article 17.
- Lewinger, J. and Bull, S. (2006) Validity, Efficiency, and Robustness of Family-Based Test of Association. *Genetic Epidemiology* 30, 62-76.
- Lunetta, K., Farove, S., Biederman, J., and Laird, N. (2000) Family-based tests of association and linkage using unaffected sibs, covariates and interaction. *American Journal of Human Genetics* 66, 605-614.
- Lyon, H., Emilsson, V., Hinney, A., Heid, I., Lasky-Su, J., Zhu, X., Thorleifsson, G., Gunnarsdottir, S., Walters, G., Thorsteinsdottir, U., Kong, A., Gulcher, J., Nguyen, T., Scherag, A., Pfeufer, A., Meitinger, T., Bronner, G., Rief, W., Soto-Quiros, M., Avila, L., Klanderma, B., Raby, B., Silverman, E., Weiss, S., Laird, N., Ding, X., Groop, L., Tuomi, T., Isomaa, B., Bengtsson, K., Butler, J., Cooper, R., Fox, C., O'Donnell, C., Vollmert, C., Celedon J., Wichmann, H., Hebebrand, J., Stefansson, K., Lange, C., and Hirschhorn, J. (2007) The association of a SNP upstream of *INSIG2* with Body Mass Index is reproduced in several but not all cohorts. *PLoS Genetics* 3(4).
- Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N., and Reich, D. (2006) Principal components analysis corrects for stratification

- in genome-wide association studies. *Nature Genetics* 38(8), 904-9.
- Pritchard, J. and Donnelly, P. (2001) Case-Control studies of association in structured or admixed population. *Theoretical Population Biology* 60, 227-237.
- Rabinowitz, D., and Laird, N. (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.
- Risch, N. (2000) Searching for genetic determinants in the new millennium. *Nature* 405, 847-856.
- Risch, N., Burchard, E., Ziv, E., and Tang, H. (2002) Categorization of humans in biomedical research: genes, race and disease. *Genome Biology* July 1, 3(7).