Statistical inference of biometrical genetic model with cultural transmission

Xiaobo Guo, Tian Jin, Xueqin Wang, Heping Zhang* and Shouqiang Zhong

Twin and family studies establish the foundation for studying the genetic, environmental and cultural transmission effects for phenotypes. In this work, we make use of the well established statistical methods and theory for mixed models to assess cultural transmission in twin and family studies. Specifically, we address two critical yet poorly understood issues: the model identifiability in assessing cultural transmission for twin and family data and the biases in the estimates when sub-models are used. We apply our models and theory to two real data sets. A simulation is conducted to verify the bias in the estimates of genetic effects when the working model is a sub-model.

Keywords and phrases: Twin and family study, Biometrical genetic model, Cultural transmission, Biometrical genetic model, Identifiability, Likelihood ratio test, Mixed-effects model.

1. INTRODUCTION

Twin and family study designs are often used to assess the heritability of a phenotype by partitioning the genetic and environmental contributions to the phenotype. One of the most popular approaches to analyzing such data is structural equation modeling (SEM). Meanwhile, several software packages are available for performing SEM including MX [21, 23], LISREL [15, 22] and Mplus [20]. Despite the popularity of these software packages, they have not incorporated advanced statistical methodologies such as nonparametric modeling and variable selection. As a result, general linear models such as mixed-effect models have emerged as an alternative for analyzing twin and family data [6, 8, 19, 24, 25, 28].

Recently, [28] introduced a comprehensive approach for the analysis of twin and family data and addressed some fundamental questions such as identifiability and the asymptotic properties of the likelihood ratio statistic under the mixed model framework. An underlying assumption of the genetic models discussed in [28] is that the resemblance between parents and offspring is determined by genes. However, in addition to genetic transmission, the resemblance

*Corresponding author.

between parents and offspring may also result from cultural transmission [2]. In the case of cultural transmission, resemblance between parents and offspring is often caused by the effects of parents' phenotypes on offspring behaviors or phenotypes. In the classical twin design, cultural transmission is regarded as part of common environmental effects. An extended twin design such as parent-twin design [1, 4, 7, 14, 16, 18] provides a solution to partition the offspring common environmental effects into the cultural transmission and the shared environmental effects among offspring only. To our knowledge, the existing methods use structural equation modeling for the extended twin design with cultural transmission. The established theory and methods for mixed models have not been applied to deal with parent-twin data. Our first aim here is to fill this gap, and present a method to infer the cultural transmission in the extended twin design.

While it is important to model cultural transmission in parent-twin data, the identifiability problem makes it difficult to dissect the additive genetic effects (A), dominant genetic effects (D), common environmental effect (C), unique environmental effect (E), and cultural transmission simultaneously. Thus, a common strategy is to consider sub-models by, for example, ignoring the dominant genetic effect, say the ACE-fm model; or deleting the cultural transmission, that is the ACDE model. The theory about the ACDE model has been well established [28]. However, there is no established theory for the ACE-fm model, even though it is widely used in the twin study. Moreover, the estimates would be biased under the ACE-fm. Our second aim is to investigate the identifiability problem for the ACE-fm model and assess the robustness of the estimates when the fitted model is different from the data-generating model.

The rest of the paper is organized as follow. In Section 2, we propose a general mixed model, which includes parents' transmission effects, for parent-twin quartet data. The estimates of the sub-model are investigated in Section 3. In Section 4, we use the proposed approach to analyze two real datasets. A brief discussion is presented in Section 5.

2. MODEL WITH CULTURAL TRANSMISSION

In genetic models, we decompose the total variance of the trait into four components: additive (A) and dominant

(D) genetic effects, common environmental effects (C), and random error (E) [5]. Specifically, we have

$$(1) \qquad Y_{ij} = x_{ij}^T\beta + A_{ij} + C_{ij} + D_{ij} + E_{ij},$$

where Y_{ij} is the trait value of individual j in family i, $x_{ij}^T\beta$ is the systematic part with $E(Y_{ij}) = x_{ij}^T\beta$, A_{ij}, C_{ij} and D_{ij} follow normal distribution with mean 0 and variance $var(A_{ij}) = \sigma_A^2, var(C_{ij}) = \sigma_C^2$ and $var(D_{ij}) = \sigma_D^2$, representing additive genetic, common environmental and dominant genetic effects, respectively. Independent of the other components in model (1) and each other, E_{ij} is a normal random variable with mean 0 and variance σ_E^2, representing residual environmental random effects. Let $j = 1, 2, 3, 4$ refer to the father, mother, and the twins, respectively. Model (1) is commonly referred to as the ACDE model.

According to genetic theory [5], the additive genetic effects of twin pair can be decomposed as

$$(2) \qquad A_{ij} = (A_{i1} + A_{i2})/2 + \tilde{A}_{ij}, \quad j = 3, 4,$$

where A_{i1}, A_{i2} represent offsprings' additive genetic effect which are transmitted by their parents, and $\tilde{A}_{ij}$ are the residual additive genetic effects which are independent of A_{i1}, A_{i2} and have variance $\sigma_A^2/2$. Moreover, $\tilde{A}_{i3} = \tilde{A}_{i4}$ for MZ twins while for DZ twins, $\tilde{A}_{i3}$ and $\tilde{A}_{i4}$ are independent. For the dominant genetic effects, the correlation between parents and twins pairs is $corr(D_{i1(i2)}, D_{i3(i4)}) = 0$, while $corr(D_{i3}, D_{i4}) = 1/4$ for MZ twin pairs and $corr(D_{i3}, D_{i4}) = 0$ for DZ twin pairs respectively.

As for the cultural transmission effect, the resemblance between parents and offspring is often caused by the effects of parents' phenotypes on offspring's common environment. Therefore, the common environmental effects of twin pair can be dissected as follows:

$$(3) \quad C_{i3} = C_{i4} = (Y_{i1} - x_{i1}^T\beta_{i1})f + (Y_{i2} - x_{i2}^T\beta_{i2})m + \tilde{C}_{i3},$$

where $(Y_{i1} - x_{i1}^T\beta_{i1})f$ and $(Y_{i2} - x_{i2}^T\beta_{i2})m$ represent the shared environment contributed by father and mother, respectively, and $\tilde{C}_{i3}$ is the common environment established by twins' cohort with variance $\tilde{\sigma}_C^2$. We define cultural transmission rate

$$h_c = \frac{\sigma_C^2 - \tilde{\sigma}_C^2}{\sigma_C^2},$$

which indicates the proportion of the common environmental effect arising from parental cultural transmission.

According to formula (3), testing whether $\sigma_C^2 = 0$ is equivalent to test $\tilde{\sigma}_C^2 = 0, f = 0, m = 0$. However, the parameter space of $\tilde{\sigma}_C^2$ does not meet the standard regularity conditions because $\tilde{\sigma}_C^2 = 0$ lies on the boundary of the parametric space $[0, +\infty)$. According to [26], we can derive that the asymptotic distribution of the likelihood ratio statistics for testing $H_0 : f = m = \tilde{\sigma}_C^2 = 0$ is a a mixture distribution

of $\chi_2^2 : \chi_3^2$ with mixing probabilities $0.5 : 0.5$. The proof is given in the appendix.

We refer to model (1) with assumption (3) as ACDE-fm model, and shall prove later that this ACDE-fm model is not fully identifiable. In applications, we often focus on two popular submodels. The first model is the well known ACDE model which ignores cultural transmission, that is by setting $f = m = 0$ in assumption (3). The second model is an ACE-fm model which is defined by letting the dominant genetic effect be zero in the ACDE-fm model.

Unlike the ACDE model without cultural transmission [28], we need to consider the covariance between A_{ij} and C_{ij} in our model. Let S denote the covariance, and then we have

$$\begin{aligned}
S &= Cov((A_{i1} + A_{i2})/2 + \tilde{A}_{i3}, f(Y_{i1} - x_{i1}^T\beta_{i1}) \\
&\quad + m(Y_{i2} - x_{i2}^T\beta_{i2}) + \tilde{C}_{i3}) \\
&= \frac{f}{2}Cov(A_{i1}, A_{i1} + C_{i1} + D_{i1} + E_{i1}) \\
&\quad + \frac{m}{2}Cov(A_{i2}, A_{i2} + C_{i2} + D_{i2} + E_{i2}) \\
&= \frac{f}{2}(\sigma_a^2 + S) + \frac{m}{2}(\sigma_a^2 + S),
\end{aligned}$$

therefore,

$$S = \frac{f + m}{2 - f - m}\sigma_A^2.$$

Note that A_{ij} and C_{ij} are not correlated only when $f + m = 0$; that is, the cultural transmissions from father and mother cancel each other out. Furthermore, we can show that there is an underlying constraint for cultural transmissions; that is $f + m < 2$, and S has the same sign with $f + m$. The detail is given in Appendix A.1.

Let $\sigma_{jk} = cov(Y_{ij}, Y_{ik})$, where $j, k \in \{1, 2, 3, 4\}$. By simple calculations, we can obtain

$$\sigma_{jj} = \sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2 + 2S, \quad j = 1, 2, 3, 4,$$
$$\sigma_{13} = \sigma_{14} = f(\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2 + 2S) + (\sigma_A^2 + S)/2,$$
$$\sigma_{23} = \sigma_{24} = m(\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2 + 2S) + (\sigma_A^2 + S)/2.$$

For MZ twins,

$$\sigma_{34} = \sigma_A^2 + \sigma_C^2 + \sigma_D^2 + 2S.$$

For DZ twins,

$$\sigma_{34} = \frac{1}{2}\sigma_A^2 + \sigma_C^2 + \frac{1}{4}\sigma_D^2 + 2S.$$

Let h denote the broad sense heritability, namely, the ratio of genetic variation to the total phenotypic variation. We have

$$h = \frac{\sigma_A^2 + \sigma_D^2}{Var(Y)} = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2 + 2S}.$$
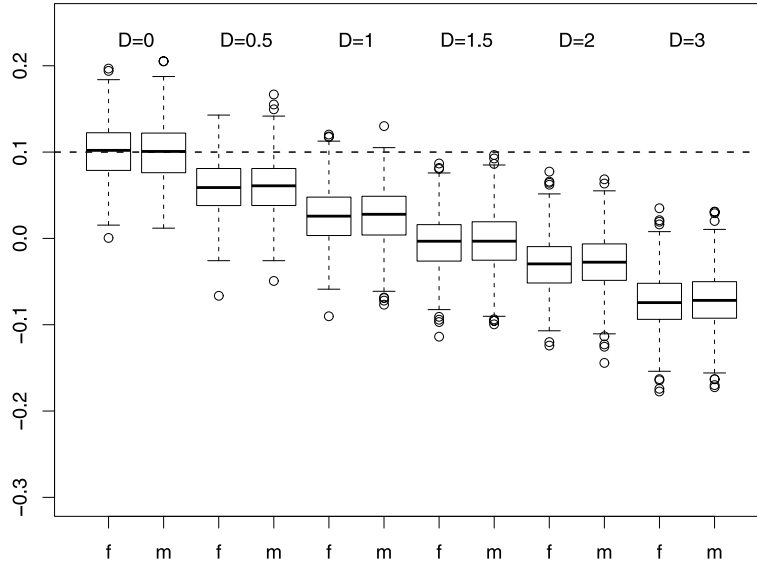
Figure 1. The estimates of cultural transmissions, f and m, obtained in the ACE-fm models with the true value of the dominant genetic effect ranged from 0 to 3. The dash lines represent the true values of f and m.

Unfortunately, an identifiability problem arises when fitting ACDE-fm model for parent-twin data. The following theorem helps us understand the identifiability problem.

Theorem 2.1. *Consider twin-parent quartet data. Suppose that there are n_{MZ} pairs of MZ twins and n_{DZ} pairs of DZ twins and the following conditions hold,*

1. *$n_{MZ} > 0$ and $n_{DZ} > 0$.*
2. *Random mating.*
3. *Genetic and cultural transmission is the same between son and daughter.*

Then, the ACDE-fm model is not identifiable while the ACE-fm model and ACDE model are identifiable.

The proof of this theorem is given in the Appendix. This theorem tells us the full model is not identifiable and the ACDE model as well as the ACE-fm model is identifiable. There are two ways to deal with the identifiability problem. One is to collect extended families (e.g., including cousins or uncles) so that we can have at least one additional equation between the covariance and parameters. Another way is to estimate parameters using submodels ignoring some effects. In this paper we focus on the latter solution and investigate the bias of the estimates when the data are generated from the full model.

3. THE ACE-FM MODELS

3.1 Estimation procedure

For the parent-twin data, even though we cannot identify the full ACDE-fm model directly, it is meaningful to assess the robustness of the estimates when the fitted model is different from the data-generated model. In this section, we investigate the parameter estimates from the ACE-fm models

when the data were *generated* from an additive genetic effect, dominant genetic effect, common environmental effect, unique environmental genetic effect, and cultural transmission effect, respectively. For clarity, let $\lambda_A^2, \lambda_C^2, \lambda_D^2, \lambda_E^2, \lambda_f$, and λ_m denote the variances of the random effects A, C, D, E, paternal and maternal cultural transmission effects in the working genetic models, respectively.

Theorem 3.1. *Suppose that $\tilde{\lambda}_A^2, \tilde{\lambda}_C^2, \tilde{\lambda}_E^2, \tilde{\lambda}_f$ and $\tilde{\lambda}_m$ are the maximum likelihood estimators obtained under the ACD-fm model with parent-twin data. When dominant genetic effect exists, $\tilde{\lambda}_E^2$ is consistent while $\tilde{\lambda}_A^2, \tilde{\lambda}_C^2, \tilde{\lambda}_f$ and $\tilde{\lambda}_m$ are asymptotically biased. Specifically, we have $\tilde{\lambda}_A^2 \xrightarrow{P} \sigma_A^2 + \frac{3}{2}\sigma_D^2$.*

Theorem 3.1 tells us $\tilde{\sigma}_A^2$ estimated by ACE-fm model is overestimated when $\sigma_D^2 > 0$ and the bias might be relatively large when σ_D^2 is large in reference to σ_A^2. The bias in the estimates of σ_C, f and m are more complex, and hence is examined through simulations.

3.2 Simulation

In this section we perform simulation studies to verify the results presented in Section 3.1. We generated 1,000 data sets. Each data set consists of 500 families with MZ twin pairs and 500 families with DZ twin pairs. Without loss of generality, we set the true values as follows: $\sigma_A^2 = \sigma_C^2 = 3$, $\sigma_E^2 = 1$, $f = m = 0.1$ and let σ_D^2 be 0, 0.5, 1, 1.5, 2 and 3, respectively. With the data simulated above, the ACE-fm is fitted. The estimates of parental cultural transmissions f and m obtained in the ACE-fm model are presented in Figure 1, while Figure 2 shows the comparisons of the heritability estimates under the ACE-fm and the true heritabilities.
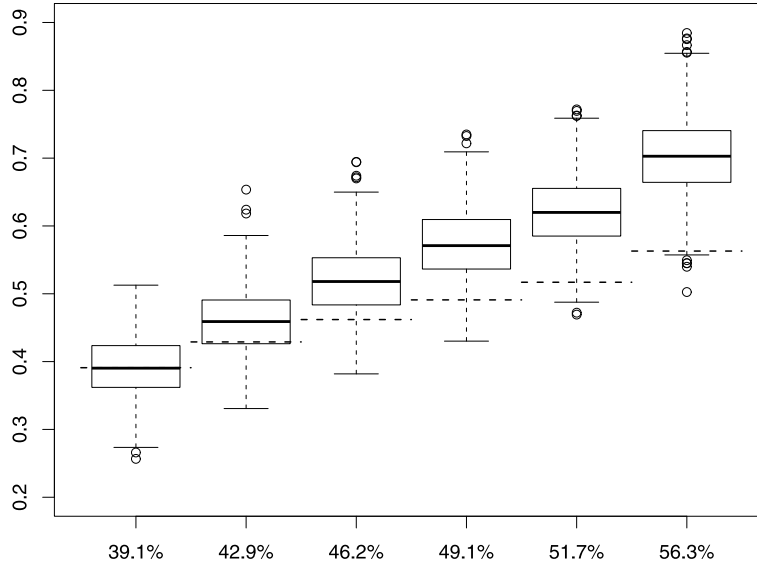
Figure 2. The estimates of heritability obtained from the ACE-fm with the true value of the dominant genetic effect ranging from 0 to 3. The dash lines represent the true values of heritability, which are also illustrated in the abscissa.

It is evident from Figure 1 that the estimates of f and m are underestimated. In particular, when the dominant genetic effect are relatively large, such as 2 or 3 in our simulations, the estimates of f and m can be negative though the true values of f and m are 0.1.

In terms of heritability estimates, we can see from Figure 2 that the heritability estimate is overestimated and the bias increases as the true value of the dominant genetic effect rises, which confirms the theoretical result in Theorem 3.1.

4. APPLICATION

4.1 Estimating the cultural transmission of Anterior Chamber Depth (ACD)

ACD, which is considered as an ideal intermediate phenotype for angle closure, has been recognized as the cardinal anatomic risk factor for angle closure. The data are from Guangzhou Twin Eye Study Center [12] which consists of 563 families, 2,058 individuals, 411 fathers, 521 mothers, and 563 twins (357 MZ twins and 206 DZ twins).

In [28], the ACDE model was fitted for this dataset and an insignificant dominant genetic effect was found, which suggests that the dominant genetic effect is relatively small. Here we apply the ACE-fm model, which sets the dominant genetic effect to 0, to re-analyze the ACD data. We include age, age × age, and sex as covariates. *There were 194 parents who did not participate in this study, and hence are not considered in the analysis.* The estimates are computed using Matlab function "fminsearch", which uses the simplex search method [17]. Note that the p-value for λ_C^2 was obtained from the mixture distribution $\frac{1}{2}\chi_3^2 + \frac{1}{2}\chi_2^2$, whereas the p-values for the remaining variance components were derived from the mixture distribution $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ [28].

Table 1. The estimates for the ACD and AOD data based on the ACE-fm model

Parameter	ACD		AOD	
	Estimated	p-value	Estimated	p-value
λ_A^2	0.0587	$<.0001$	0.0329	$<.0001$
λ_C^2	0.0178	$<.0001$	0.0076	0.0021
λ_E^2	0.0060	$<.0001$	0.0144	$<.0001$
f	0.0108	0.8467	-0.1493	0.026
m	-0.0855	0.0971	-0.138	0.0316
Intercept	3.2366	$<.0001$	0.5780	$<.0001$
Age	0.0211	$<.0001$	0.0103	$<.0001$
Age^2	-0.0005	$<.0001$	-0.0003	$<.0001$
Sex	0.1117	$<.0001$	0.0491	$<.0035$
Heritability	75.0%		70.5%	

From Table 1, we can see that a significant additive genetic effect and common environment effect are detected. In the ACE-fm model, no significant cultural transmission effects are detected. The estimated heritability is 75.0% using the ACE-fm model.

4.2 Estimating the cultural transmission of Angle Opening Distance (AOD)

Population-based studies suggest that the prevalence of primary angle-closure glaucoma (PAGG) is higher in Chinese than European and African populations [9, 10]. Previous cross-sectional studies have demonstrated that the persons with narrow drainage angles have a higher risk for the development of PAC-related problems [11]. Here, angle width is represented by the AOD, as well as the angle recess area (ARA) and the trabecular-iris space area (TISA).The AOD data are from Guangzhou Twin Eye Study Center [13]

which include 476 families: 276 fathers, 400 mothers, and 476 twins (311 MZ twins and 166 DZ twins). Here, the ACE-fm model was fitted *and there were 284 parents who did not participate in this study*. The same asymptotic distributions for the likelihood ratio test statistics as those used in Section 4.1 are applied to derive the p values for the tests of random effects. The estimates of the ACE-fm are presented in the right-hand s Table 1.

From Table 1, a significant additive genetic effect and common environmental effect are detected in the ACE-fm model. Meanwhile, two negative and significant cultural transmission effects are revealed. The estimated heritability is 70.5%. On the other hand, according to Section 2, we can obtain that the cultural transmission effect explains 4.2% of the total phenotype variation, and the correlation between the common environmental and additive genetic effects is -0.26. Meanwhile, the cultural transmission rate is 25.4%.

In [28], the ACDE model was fitted to this data and the additive and dominant genetic effects are $0.0149(pvalue < 0.0001)$ and $0.0146(pvalue = 0.0007)$, both of them are extremely significant. From Figure 1, we can see that if the dominant genetic effect is relatively large, the bias of the estimates of cultural transmission effects would become severe. In this data analysis, the negative and significant cultural transmission effect could be possibly caused by the neglected dominant genetic effect and a further research about the cultural transmission problem is needed.

In summary, considering the dominant genetic effect is commonly seen in practice, in the real data analysis, a reasonable approach is to fit the ACDE model firstly. In the second stage, if the dominant genetic effect is relatively small, we neglect the dominant genetic effect and turn to fit the ACE-fm model. However, if the dominant genetic effect cannot be ignored, we could apply the CDE-fm or ACDE model depending on the relative effect size of the dominant effect. Therefore, according to the two stage strategy, the ACDE model is preferable for the AOD data. It should be noted that biases still exist in the proposed two stage approach since it is impossible to avoid the bias issue due to the lack of identifiability for the full model.

5. DISCUSSION

In this article, we illustrate how to analyze complex model of twin and family data in the mixed model framework. An advantage of the mixed models is that the general methodology and theory are well established. In addition, it is straightforward to accommodate the covariates, consider nonparametric trends of the covariates, and incorporate variable selection strategy. Taking advantages of the well established statistical theory for mixed models, we made two contributions to the understanding of cultural transmission with parent-twin data. First, we characterized the conditions for the identifiability for the ACE-fm models. Second, we derived the asymptotic distributions of the

likelihood ratio test statistics for testing the common environment effect. We should note that the naive χ_3^2 would produce a conservative p-value [3, 27, 28].

We illustrated the bias problem when using ACE-fm model for twin-family data through both theory and simulation. Specifically, the estimates of cultural transmission in ACE-fm model is always underestimated and the bias becomes more severe when the dominant genetic effect is relatively large. If the dominant genetic effect plays an important role in the total variation, the analysis based on the ACE-fm model can lead to a misleading conclusion for the existence of the cultural transmission. This problem can be overcome if the families can be extended to include more relatives, especially after the data were already collected. However, in reality, this may not be feasible. Due to the lack of identifiability, we have to make a trade off, and decide on a case by case basis. In the simulation section, we proposed a reasonable two stage approach. We first fitted the ACDE model to the data. In the next step, according the relative effect sizes of additive and dominant genetic effect obtained in the first step, we would apply the ACE-fm, CDE-fm or ACDE to the data and get an updated result for the data. *One of the anonymous referees suggested that we first estimate the dominant effect with the ACDE model and then estimate the cultural transmission effect in the ACDE-fm model by plugging in the estimated dominant effect from the first stage. This approach is appealing due to its control of type I error, but its power may be compromised. A thorough evaluation of this approach warrants further investigation.*

In our real data analysis part, even we detected a significant cultural transmission effect for the AOD data. Due to the large amount of dominant genetic effects obtained in the ACDE model, this result is unreliable and more research is needed.To our best knowledge, despite the perceived influence of culture, it is rare to observe a large amount of cultural transmission effect [1, 2, 4, 7, 14, 16, 18]. There could be a number of reasons. Firstly, the parents and children are measured at different ages, resulting in changed cultures. Secondly, the characteristics in parents may not directly affect the same characteristics in children or parents may pass only their environmental aspect to their offspring [2]. Thirdly, the children may be more sensitive to their peers than to their parents. Finally, from the theoretical point, the ignored dominant effect may play an important role in the phenotype, which reduces power in detecting the cultural transmission effect. In any case, this issue is not well understood.

Appendix 1: Proof of the asymptotic distribution of the likelihood ratio statistic for testing lies on the boundary of the parametric space $[0, +\infty)$

According to [26], the asymptotic distribution of the likelihood ratio statistics for testing $H_0 : f = m = \tilde{\sigma}_C^2 = 0$ is

the equivalent to the distribution of

$$U = \inf_{\theta \in \Omega_0} \| Z - \theta \|^2 - \inf_{\theta \in \Omega_1} \| Z - \theta \|^2$$

where $\Omega_0 = (0,0,0)$ represents the parameter space under null hypothesis $H_0 : f = m = \tilde{\sigma}_C^2 = 0$, $\Omega_1 = (0, +\infty) \times R^2$ denotes as parameter space under the alternative hypothesis and $Z = (Z_1, Z_2, Z_3)' \sim N(0, I_3)$. The formula of U can be simplified as

$$U = \inf_{\theta \in \Omega_0} \| Z - \theta \|^2 - \inf_{\theta \in \Omega_1} \| Z - \theta \|^2$$
$$= Z_1^2 + Z_2^2 + Z_3^2 - \inf_{\theta_1 \in (0, +\infty)} (Z_1 - \theta_1)^2.$$

When $Z_1 \geq 0$, U reduces to $Z_1^2 + Z_2^2 + Z_3^2$, which follows χ_3^2. However, if $Z_1 \leq 0$, U is equal to $Z_2^2 + Z_3^2$, which follows χ_2^2. Therefore, U follows a mixture distribution of $\chi_2^2 : \chi_3^2$ with mixing probabilities $0.5 : 0.5$.

Appendix 2: Proof of $f + m \leq 2$

For the ACE-fm model, firstly, according to the Cauchy-Schwartz inequality, we have

$$S^2 = (cov(A,C))^2 \leq var(A)var(C) = \sigma_A^2 \sigma_C^2.$$

On the other hand, depending on equation (3), it follows

$$\sigma_C^2 = (f^2 + m^2)\sigma_{R_p}^2 + \tilde{\sigma}_C^2$$
$$= (\sigma_A^2 + \sigma_C^2 + \sigma_E^2 + 2S)(f^2 + m^2) + \tilde{\sigma}_C^2.$$

Assume $f + m > 2$.
Since $S = \frac{f+m}{2-f-m}\sigma_A^2$ together with $S^2 \leq \sigma_A^2 \sigma_C^2$, we obtain $\sigma_A^2 = (\frac{2-f-m}{f+m})S$ and $S \geq (\frac{2}{f+m} - 1)\sigma_C^2$; therefore,

$$\sigma_C^2 \geq (\sigma_A^2 + \sigma_C^2 + \sigma_E^2 + 2S)(f^2 + m^2)$$
$$= \left\{ \left(\frac{2}{f+m} + 1 \right) S + \sigma_C^2 + \sigma_E^2 \right\} (f^2 + m^2)$$
$$\geq \left\{ \left(\frac{2}{f+m} + 1 \right) \left(\frac{2}{f+m} - 1 \right) \sigma_C^2 + \sigma_C^2 \right\} (f^2 + m^2)$$
$$= \frac{4(f^2 + m^2)}{(f+m)^2} \sigma_C^2$$
$$> 2\sigma_C^2,$$

which is a contradiction. In conclusion, we obtain $f + m \leq 2$.

Proof of Theorem 2.1

A genetic model is identifiable if and only if the covariance matrix of $\mathbf{Y}$ is identifiable, since the phenotype $\mathbf{Y}$ follows a multivariate normal distribution. For twin-parent quartet data with both MZ twins and DZ twins, there are 5 distinct elements in the covariance matrix: $V_1(\theta) = \sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2 + 2S$, $V_2(\theta) = f(\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2 + 2S) + (\sigma_A^2 + S)/2$, $V_3(\theta) = m(\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2 + 2S) + (\sigma_A^2 + S)/2$, $V_4(\theta) =$

$\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + 2S$ and $V_5(\theta) = \frac{1}{2}\sigma_A^2 + \sigma_C^2 + \frac{1}{4}\sigma_D^2 + 2S$, where θ is a vector of parameters. Therefore, a model is nonidentifiable if there are more than 5 parameters in the covariance matrixes, implying that the ACDE-fm model is nonidentifiable.

Furthermore, for ACE-fm model with $\theta = (\sigma_A^2, \sigma_C^2, \sigma_E^2, f, m)^T$ and $\sigma_D^2 = 0$, we can show that $(V_1(\theta_1), V_2(\theta_1), V_3(\theta_1), V_4(\theta_1), V_5(\theta_1))^T = (V_1(\theta_2), V_2(\theta_2), V_3(\theta_2), V_4(\theta_2), V_5(\theta_2))^T$ is equivalent to $\theta_1 = \theta_2$, and thus the ACE-fm model is identifiable. To confirm this statement, first we note that $f_i + m_i \leq 2$ where $i = 1, 2$. Then we solve the equations

$$\sigma_{RP1}^2 = \sigma_{RP2}^2, \tag{4}$$

$$f_1 \sigma_{RP1}^2 + (\sigma_{A1}^2 + S_1)/2 = f_2 \sigma_{RP2}^2 + (\sigma_{A2}^2 + S_2)/2, \tag{5}$$

$$m_1 \sigma_{RP1}^2 + (\sigma_{A1}^2 + S_1)/2 = m_2 \sigma_{RP2}^2 + (\sigma_{A2}^2 + S_2)/2, \tag{6}$$

$$\sigma_{A1}^2 + \sigma_{C1}^2 + 2S_1 = \sigma_{A2}^2 + \sigma_{C2}^2 + 2S_2, \tag{7}$$

$$\frac{1}{2}\sigma_{A1}^2 + \sigma_{C1}^2 + 2S_1 = \frac{1}{2}\sigma_{A2}^2 + \sigma_{C2}^2 + 2S_2, \tag{8}$$

with $\sigma_{RPi}^2 = (\sigma_{Ai}^2 + \sigma_{Ci}^2 + \sigma_{Ei}^2 + 2S_i)$ and $S_i = (f_i + m_i)\sigma_{Ai}^2/(2 - f_i - m_i)$, $i = 1, 2$. Combining (4), (7) and (8), we can get

$$\sigma_{A1}^2 = \sigma_{A2}^2 \quad \text{and} \quad \sigma_{E1}^2 = \sigma_{E2}^2.$$

Based on (6) and (7) and after some calculations, we have

$$\sigma_{RP1}^2 (f_1 + m_1 - f_2 - m_2) \tag{9}$$
$$+ \sigma_{A1}^2 \left(\frac{f_1 + m_1}{2 - f_1 - m_1} - \frac{f_2 + m_2}{2 - f_2 - m_2} \right) = 0.$$

A trivial solution to (4)–(8) is $\theta_1 = \theta_2$. The other solution is

$$f_1 + m_1 = 2 + \frac{2\sigma_{A1}^2}{\sigma_{RP1}^2 (2 - f_2 - m_2)},$$

which is larger than 2 and contradicts with the necessary condition for the ACE-fm model, i.e., $f_i + m_i \leq 2$. Hence the proof is completed.

Proof of Theorem 3.1

Under the ACE-fm model, the log-likelihood is

$$l(\lambda) = -(n_{MZ} + n_{DZ})\log(2\pi) - \frac{n_{MZ}}{2}\log|\Sigma_{MZ}(\lambda)|$$
$$- \Sigma_{MZpairs}\left(\frac{1}{2}(y_i - \mu)'\Sigma_{MZ}^{-1}(\lambda)(y_i - \mu) \right)$$
$$- \frac{n_{DZ}}{2}\log|\Sigma_{DZ}(\lambda)|$$
$$- \Sigma_{DZpairs}\left(\frac{1}{2}(z_i - \mu)'\Sigma_{DZ}^{-1}(\lambda)(z_i - \mu) \right),$$

where

$$\Sigma_{MZ}(\lambda) = \begin{bmatrix} \lambda_{rp} & 0 & \lambda_{ft} & \lambda_{ft} \\ 0 & \lambda_{rp} & \lambda_{mt} & \lambda_{mt} \\ \lambda_{ft} & \lambda_{mt} & \lambda_{rp} & \lambda_{MZ} \\ \lambda_{ft} & \lambda_{mt} & \lambda_{MZ} & \lambda_{rp} \end{bmatrix}$$

and

$$\Sigma_{DZ}(\lambda) = \begin{bmatrix} \lambda_{rp} & 0 & \lambda_{ft} & \lambda_{ft} \\ 0 & \lambda_{rp} & \lambda_{mt} & \lambda_{mt} \\ \lambda_{ft} & \lambda_{mt} & \lambda_{rp} & \lambda_{DZ} \\ \lambda_{ft} & \lambda_{mt} & \lambda_{DZ} & \lambda_{rp} \end{bmatrix}$$

with $\lambda_{rp} = \lambda_A^2 + \lambda_C^2 + \lambda_E^2 + 2S$, $\lambda_{ft} = \frac{2\lambda_f\lambda_{rp}+\lambda_A^2+S}{2}$, $\lambda_{mt} = \frac{2\lambda_m\lambda_{rp}+\lambda_A^2+S}{2}$, $\lambda_{MZ} = \lambda_A^2 + \lambda_C^2 + 2S$, $\lambda_{DZ} = \frac{1}{2}\lambda_A^2 + \lambda_C^2 + 2S$ and $S = (\lambda_f + \lambda_m)\lambda_A^2/(2 - \lambda_f - \lambda_m)$.

The maximum likelihood estimations of $\lambda_A^2, \lambda_C^2, \lambda_E^2, \lambda_f$ and λ_m are unbiased estimates from the solutions of equations $E[\frac{\partial l(\lambda)}{\partial \lambda_A^2}] = 0, E[\frac{\partial l(\lambda)}{\partial \lambda_C^2}] = 0, E[\frac{\partial l(\lambda)}{\partial \lambda_E^2}] = 0, E[\frac{\partial l(\lambda)}{\partial \lambda_f}] = 0$ and $E[\frac{\partial l(\lambda)}{\partial \lambda_m}] = 0$. It is easy to obtain that $\tilde{\lambda}_E^2 = \sigma_E^2$, $\tilde{\lambda}_A^2 = \sigma_A^2 + 3\sigma_D^2/2$. Therefore, we have $\tilde{\lambda}_E^2 \xrightarrow{P} \sigma_E^2$ and $\tilde{\lambda}_A^2 \xrightarrow{P} \sigma_A^2 + \frac{3}{2}\sigma_D^2$. By the same reasoning, $\tilde{\lambda}_C^2$, $\tilde{\lambda}_f$ and $\tilde{\lambda}_m$ can be calculated, but the expressions are very complicated.

ACKNOWLEDGEMENT

REFERENCES

[1] BOOMSMA, D. and MOLENAAR, P. (1987). Constrained maximum likelihood analysis of familial resemblance of twins and their parents. *Acta Genet Med Gemellol* **36** 29–39.

[2] DE MOOR, M. H. M., WILLEMSEN, G., REBOLLO-MESA, I., STUBBE, J. H., DE GEUS, E. J. C. and BOOMSMA, D. I. (2011). Exercise participation in adolescents and their parents: Evidence for genetic and generation specific environmental effects. *Behavior Genetics* **41** 211–222.

[3] DOMINICUS, A., SKRONDAL, A., GJESSING, H. K., PEDERSEN, N. L. and PALMGREN, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics* **36** 331–340.

[4] EAVES, L. J., LAST, K. A., YOUNG, P. A. and MARTIN, N. G. (1978). Model-fitting approaches to the analysis of human behaviour. *Heredity* **41** 249–320.

[5] FALCONER, D. and MACKAY, T. (1996). *Introduction to Quantitative Genetics.* Longman.

[6] FENG, R., ZHOU, G., ZHANG, M. and ZHANG, H. (2009). Analysis of twin data using sas. *Biometrics* **65** 584–589. MR2751483

[7] FULKER, D. (1982). Extensions of the classical twin method. *Progress in Clinical and Biological Research* **103** 395.

[8] GUO, G. and WANG, J. (2002). The mixed or multilevel model for behavior genetic analysis. *Behavior Genetics* **32** 37–49.

[9] HE, M., FOSTER, P., JOHNSON, G. and KHAW, P. (2005). Angle-closure glaucoma in East Asian and European people. Different diseases? *Eye* **20** 3–12.

[10] HE, M., FOSTER, P. J., GE, J., HUANG, W., ZHENG, Y., FRIEDMAN, D. S., LEE, P. S. and KHAW, P. T. (2006a). Prevalence and clinical characteristics of glaucoma in adult Chinese: A population-based study in Liwan District, Guangzhou. *Investigative Ophthalmology & Visual Science* **47** 2782.

[11] HE, M., FOSTER, P. J., GE, J., HUANG, W., WANG, D., FRIEDMAN, D. S. and KHAW, P. T. (2006b). Gonioscopy in adult Chinese: The Liwan eye study. *Investigative Ophthalmology & Visual Science* **47** 4772.

[12] HE, M., WANG, D., ZHENG, Y., ZHANG, J., YIN, Q., HUANG, W., MACKEY, D. A. and FOSTER, P. J. (2008a). Heritability of anterior chamber depth as an intermediate phenotype of angle-closure in Chinese: The Guangzhou Twin Eye Study. *Investigative Ophthalmology & Visual Science* **49** 81.

[13] HE, M., GE, J., WANG, D., ZHANG, J., HEWITT, A. W., HUR, Y. M., MACKEY, D. A. and FOSTER, P. J. (2008b). Heritability of the iridotrabecular angle width measured by optical coherence tomography in Chinese children: The Guangzhou twin eye study. *Investigative Ophthalmology & Visual Science* **49** 1356.

[14] HEATH, A. C., KENDLER, K. S., EAVES, L. J. and MARKELL, D. (1985). The resolution of cultural and biological inheritance: Informativeness of different relationships. *Behavior Genetics* **15** 439–465.

[15] JÖRESKOG, K. G. and SÖRBOM, D. (1986). *LISREL VI, Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods.* Scientific Software, Inc. Mooresville, IN.

[16] KELLER, M. C., MEDLAND, S. E., DUNCAN, L. E., HATEMI, P. K., NEALE, M. C., MAES, H. H. M. and EAVES, L. J. (2009). Modeling extended twin family data I: Description of the Cascade model. *Twin Research and Human Genetics* **12** 8–18.

[17] LAGARIAS, J. C., REEDS, J. A., WRIGHT, M. H. and WRIGHT, P. E. Convergence properties of the Nelder-Mead simplex method in low dimensions.

[18] MAES, H. H., NEALE, M. C., MEDLAND, S. E., KELLER, M. C., MARTIN, N. G., HEATH, A. C. and EAVES, L. J. (2009). Flexible Mx specification of various extended twin kinship designs. *Twin Research and Human Genetics* **12** 26–34.

[19] MCARDLE, J. J. and PRESCOTT, C. A. (2005). Mixed-effects variance components models for biometric family analyses. *Behavior Genetics* **35** 631–652.

[20] MUTHÉN, L. K. and MUTHÉN, B. O. (1998). Mplus users guide.

[21] NEALE, M. C. and CARDON, L. R. (1992). *Methodology for Genetic Studies of Twins and Families.* Springer, The Netherlands.

[22] NEALE, M., HEATH, A., HEWITT, J., EAVES, L. and FULKER, D. (1989). Fitting genetic models with LISREL: Hypothesis testing. *Behavior Genetics* **19** 37–49.

[23] NEALE, M. C., (1997). *Mx: Statistical Modeling.* Department of Psychiatry, Medical College of Virginia.

[24] PAWITAN, Y., REILLY, M., NILSSON, E., CNATTINGIUS, S. and LICHTENSTEIN, P. (2004). Estimation of genetic and environmental factors for binary traits using family data. *Statistics in Medicine* **23** 449–465.

[25] RABE-HESKETH, S., SKRONDAL, A. and GJESSING, H. (2008). Biometrical modeling of twin and family data using standard mixed model software. *Biometrics* **64** 280–288. MR2422844

[26] SELF, S. G. and LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82** 605–610. MR0898365

[27] VISSCHER, P. M. (2006). A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Research and Human Genetics* **9** 490–495.

[28] WANG, X., GUO, X., HE, M. and ZHANG, H. (2011). Statistical inference in mixed models and analysis of twin and family data. *Biometrics* **67** 987–995. MR2829273

Xiaobo Guo
Department of Statistical Science School of Mathematics and
    Computational Science
Sun Yat-Sen University
Guangzhou, 510275
China
E-mail address: mc03gxb@126.com

Tian Jin
School of Statistics and Management
Shanghai University of Finance and Economic
Shanghai, 200433
China
E-mail address: jintian82@gmail.com

Xueqin Wang
Department of Statistical Science School of Mathematics and
    Computational Science &
    State Key Laboratory of Ophthalmology
Zhongshan Ophthalmic Center &
    Zhongshan Medical School &
    Xinhua College Sun Yat-Sen University
Guangzhou, 510275
China
E-mail address: wangxq88@mail.sysu.edu.cn

Heping Zhang
Department of Epidemiology and Public Health
Yale University School of Medicine
New Haven, CT 06520-8034
USA
E-mail address: heping.zhang@yale.edu

Shouqiang Zhong
Maoming People's Hospital
Maoming 525000
China
E-mail address: mmrmyy@163.com