Phenotypic Factor Analysis of Family Data: Correction of the Bias Due to Dependency

Irene Rebollo,[1] Marleen H. M. de Moor,[1] Conor V. Dolan,[2] and Dorret I. Boomsma[1]

[1] Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands
[2] Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

Twin registries form an exceptionally rich source of information that is largely unexploited for phenotypic analyses. One obstacle to straightforward phenotypic statistical analysis is the inherent dependency, which is due to the clustering of cases within families. The present simulation study gauges the degree of the bias produced by the dependency of family data on the estimates of standard errors and chi-squared, when they are treated as independent observations in a phenotypic model, and assesses the efficiency of an estimator, which corrects for dependency. When family-clustered data are used for phenotypic analysis, in treating individuals as independent, and using standard maximum likelihood estimation, there is a tendency for the chi-square statistic to be overestimated, and the standard errors of the parameters to be underestimated. The bias increases with family resemblance, due to heritability or shared environment. The source of family resemblance — either heritability (h^2) and/or shared environment (c^2) — interacts with the composition of the sample. In the absence of c^2, samples with twins, parents and spouses show the lowest bias, whereas in the presence of c^2 samples with only twins show the lowest bias. In all conditions the bias remained below 15%. The use of the 'complex option' available in Mplus (clustering corrected robust maximum likelihood estimation) reduces the bias to the levels observed when only independent cases are considered. Thus with the use of robust estimates the bias due to family dependency becomes practically negligible in all conditions of dependency. In conclusion, the present study shows that the bias due to dependency in family data does not form a serious obstacle to phenotypic data analysis.

Twin registries form an exceptionally rich source of information due to a unique combination of characteristics. First, they generally comprise many thousands of cases. Second, the variety of measured phenotypes is large, including many psychological, biological, and clinical traits as well as important sociological and demographic information. In addition, increasingly, next of kin of twins are included (i.e., parents, siblings, spouses, children of the twins, and so on). This allows for the study of cohort effects, cultural transmission, and rater bias, and also increases the generalizability of the results to the general population. Finally, a very useful aspect of registries is that they often include longitudinal data. As Busjahn stated, 'The virtue of a twin register is not so much determined by the existing database of measures but by the ability to get back to the twins to add phenotypes in a hypothesis-driven manner' (Busjahn, 2002, p. vi). At present twin registries have been established in several countries around the world (see Boomsma, 1998; *Twin Research* Special Issue, 2002 [volume 5, number 5]), and more often than not these registries include similar or even identical phenotypic measures, which facilitates replication, and allows for cross-cultural comparisons.

Twin registries were established primarily to advance the study of genetic and environmental contributions to phenotypic individual differences. However, given the amount of information and the sample sizes, they form a rich, yet largely unexploited, source of information for phenotypic analyses (i.e., not addressing genetic or environmental sources of variance). One obstacle to straightforward phenotypic statistical analysis is the inherent dependency, which is due to the clustering of cases within families. It is generally known that simply treating dependent data as independent in phenotypic analyses results in bias in standard errors and other test (e.g., goodness-of-fit) statistics (Laplante & Hebert, 2001).

Published articles in which family data serve purely phenotypic analyses are scarce (Brown et al., 2002; Kirk et al., 1999), whereas it is quite common to apply some kind of phenotypic factor analysis (i.e., exploratory factor analysis, principal components analysis, or confirmatory factor analysis) to the measures studied, prior to the intended genetic modeling (Edwards et al.,

Irene Rebollo, Marleen H. M. de Moor, Conor V. Dolan, and Dorret I. Boomsma

1994; Eley et al., 2003; Jonnal et al., 2000; Tozzi et al., 2004; Wade et al., 2003). When confronted with the problem of nonindependence of the data, the researchers usually either ignore it, adopt complicated (yet still approximate) methods for accommodating dependency (van der Sluis et al., in press), or opt for splitting the sample and analyzing data of independent individuals only. In most cases, splitting the sample may amount to discarding as much as half of the data. This is clearly a drawback, which is exacerbated when data on multiple family members is available, since not only the size of the sample is reduced, but also the representativeness of the sample may be diminished.

The aim of the present simulation study is to gauge the degree of the bias produced by the dependency of family data on parameters, standard errors and chi-square estimates, when they are treated as independent observations in a phenotypic model, and to assess the efficiency of an estimator, which corrects for dependency. The study was designed to be representative of large and small twin registers, with and without extended family data. We consider a common factor model, with two correlated latent factors; such a model is often employed in preliminary data reduction to investigate the structure of the measures under study, and the relationship between latent constructs. We consider the effects of heritability, shared environment, family size, and estimation method on the accuracy of the parameter estimates (factor loadings, residual variances, and variance-covariance structure of the latent factors), their standard errors, and the chi-square (likelihood ratio) goodness-of-fit test. We hypothesized that (1) larger family resemblance due to heritability or shared environment will produce a decrease in the accuracy of the estimates of standard errors and chi-square; (2) when family resemblance is exclusively due to heritability, including other members of the family besides the twins, like parents or spouses, will decrease the similarity among the members of each cluster, and thus reduce the dependency and subsequent bias in the estimates of standard errors and chi-square; and (3) the use of normal theory maximum likelihood estimation (ML; Azzelini, 1996; Bollen, 1989) will lead to less accurate estimates than robust maximum likelihood estimation corrected for clustering (MLR; Muthén & Satorra, 1995).

Method

The procedure that we followed consists of two steps: (1) simulation of family (clustered) data using a common pathway model, and (2) phenotypic analysis of the simulated data, fitting a phenotypic factor model with two correlated latent factors.

Simulation of Family Data

A common pathway model (Neale & Cardon, 1992) with two common factors was used to generate family (clustered) data. Figure 1 shows the path diagram of the simulated model, for a family of dizygotic (DZ) twins, their parents and spouses.

In the simulated model, the covariance between six observed variables (V1–V6) is explained by two phenotypic correlated latent factors (F_1–F_2). The first three variables are indicators of the first factor, and the last three variables are indicators of the second factor. The values of the factor loadings and residual variances were chosen in order to have reasonable and varying signal-noise ratios: the percentage of variance explained by the latent factors was 83% for V1 and V4, 60% for V2 and V5 and 40% for V3 and V6. The same measurement model was generated for all family members: Twin 1 (T1), Twin 2 (T2), Father (F), Mother (M), and spouses of the twins (S1 and S2).

In the first set of simulations familial clustering of the data was explained by an additive genetic component (A_1 for F_1 and A_2 for F_2). The variance of each latent factor was additionally explained by a unique environmental component (E_1 for F_1 and E_2 for F_2). A genetic correlation of .5 between the two additive genetic components of the two phenotypic factors was modeled, assuming that both latent factors are correlated due to common genetic variance. Parents and offspring share 50% of the additive genetic variance, as do DZ twins on average (Falconer, 1989). Monozygotic (MZ) twins share the totality of the genetic variance, and thus the MZ model contains an additional correlation of .5 between the genetic factors of the twins $A_{1T1} \leftrightarrow A_{1T2}$ and $A_{2T1} \leftrightarrow A_{2T2}$ (not depicted in the figure), which, added to the .5 shared through the parents, makes the expected additive genetic correlation of 1.

In a second set of simulations, a general shared environmental factor 'C' was added to the variance components A and E. This factor was meant to represent environmental conditions such as socioeconomic status, or diet that might increase family resemblance across parents and their offspring, and between spouses. The general latent factor C had loadings on all the phenotypic factors F1 and F2 in all the family members.

The means of the observed variables (intercepts) as well as the means of the latent factors were fixed to zero in the simulated model. The observed variables and latent factors were multivariate normally distributed.

The simulated model includes a number of features that affect the dependency generated among family members. These features were chosen so as to resemble the empirical results found for most psychological variables studied in adult samples. Specifically, in the first set of simulations family resemblance is due to genetic factors, and effects of shared environment are absent. Under these conditions parents and twins, and the twins pairs themselves form dependent cases, whereas the spouses are mutually independent. These conditions apply to personality variables measured in adult samples (Bouchard & McGue, 2003). In the second set of simulations, the possible effects of shared environment and assortative mating were added to the model through a general shared environmental factor. Under
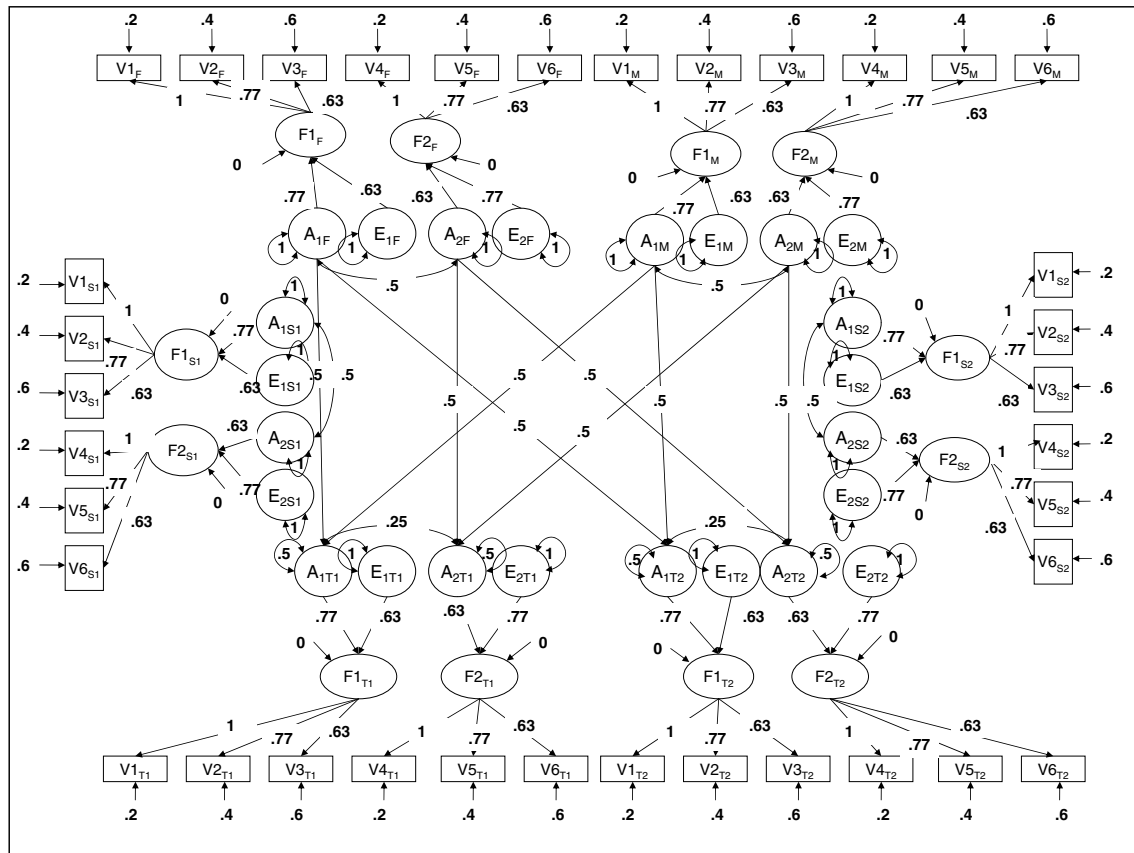
Figure 1

Simulated common pathway model.

Note: Path diagram depicted for conditions: FS = 6 (twins, parents and spouses), Heritability = 60–40. Model depicted for DZ twin families. DZ twins share 50% of their additive genetic variance inherited from their parents, as showed in the figure. MZ twins share 100% of their additive genetic variance and thus, the MZ model includes and additional correlation path of .5 across genetic factors $A_{1T1} \leftrightarrow A_{1T2}$ and $A_{2T1} \leftrightarrow A_{2T2}$. T1: Twin 1, T2: Twin 2, F: Father, M: Mother, S1: Spouse of Twin 1, S2: Spouse of Twin 2. V1–V6: observed variables, F1, F2: Common factors 1 and 2, A: Additive genetic effects, E: Nonshared environmental effects.

this condition all family members, including the spouses, form dependent cases. The results of this condition apply to cognitive abilities (Bouchard & McGue, 2003), and to social attitudes (Eaves et al., 1999).

Two factors were varied in the simulation study: (1) The degree of family resemblance through the heritability of the latent factors, and the inclusion of the C factor, and (2) family size and composition. Regarding (1), in the first set of simulations three levels of heritability were chosen: A40–A40 (.40 for both F1 and F2), A60–A40 (.60 for F1 and .40 for F2) and A60–A60 (.60 for F1 and F2). In the second set of simulations heritability was chosen to equal .35 for both factors, and the amount of variance explained by the C factor was .25 for both factors (A35–C25). In this second set of simulations the degree of resemblance between twins was equivalent to that of the A60–A60 condition, with the difference that part of it is due to shared environment, which also produces resemblance across the other family members. With regards to (2), three levels of family size (FS) were chosen: FS = 6 (twins, parents and spouses of the twins), FS = 4 (twins and parents), and FS = 2 (twins).

For the first set of simulations sample size was also varied to represent typical sample sizes of large and small twin registries (*Twin Research* Special Issue, 2002 [volume 5, number 5]). Simulations that represented large twin registries had a sample size of 2000 families (1000 MZ, 1000 DZ), and included nine conditions (three conditions of heritability × three conditions of FS). Simulations that represented small twin registries had a sample size of 500 families (250 MZ, 250 DZ), and included the three levels of heritability — family size was not varied as small twin registries often do not include other family members.

First, we carried out 1000 replications for each heritability condition with FS = 6 and a sample size of 2000. We created conditions FS = 4 and FS = 2 by selecting a subset of the generated sample, as would be done with a real data set. Second, we carried out 1000 replications for each heritability condition with FS = 2 and sample size = 500. Finally, we carried out 1000 replications for the A35–C25 condition with FS = 6 and a sample size of 2000. We created the conditions FS = 4 and FS = 2 by selecting a subset of the generated sample. We created missing data in percentages similar to those

Irene Rebollo, Marleen H. M. de Moor, Conor V. Dolan, and Dorret I. Boomsma

observed in the adult sample of the Netherlands Twin Registry (NTR) as a function of family membership: 5% for twins, 35% for parents, and 60% for spouses.

The datasets were generated[1] using a Monte Carlo procedure in Mplus Version 3.13 (Muthén & Muthén, 2005). The same model used to generate the data was fitted to the 1000 data sets using the internal Monte Carlo procedure in Mplus to ensure that the parameter values were correctly recovered. Replications that did not converge or gave inadmissible parameter estimates were excluded from further analyses.

Phenotypic Analysis of Family Data: Analysis of Bias

First, the generated data sets were restructured so that each member of the family was treated as an independent case, while retaining information about family membership. Thus, the new data sets contained six observed variables (V1–V6), and a cluster variable (family identification number). The sample size for each condition is equal to the number of families times the family size (e.g., number of families = 2000 and FS = 6 gave rise to a new sample size of 12,000).

Subsequently, we analyzed these newly created datasets in Mplus using the Monte Carlo procedure. A phenotypic factor model with two correlated latent factors was fitted using two types of estimation. The first was ML estimation, which assumes that the data are normally, identically, and independently distributed. The second type of estimation used was MLR in combination with the 'Complex' option in Mplus, which takes into account clustering of the data. The parameter estimates are ML estimates, whereas the standard errors are corrected for the dependency in the data. The correction is made by using a weight matrix that involves fourth-order moments and contains cluster information. The chi-square statistic is scale-corrected. The scale is a function of the same weight matrix and the degrees of freedom of the model (for further details about the correction, see Muthén & Satorra, 1995). The family was used as the cluster unit for the correction.

We compared the bias produced in the parameter estimates, standard errors, and chi-square statistic by the dependence of family data across heritability and FS conditions, and across the two types of estimation methods. For the first set of simulations, we expected the largest bias for the condition A60–A60 and FS = 2, and the smallest bias for the condition A40–A40 and FS = 6. For the second set of simulations, for condition A35–C25, we expected the bias for FS = 2 to be equivalent to that of the condition A60–A60. However, we expected the bias for FS = 4 or 6 to be larger than in the first set of simulations, due to the family resemblance produced by the general C. Given that the assumptions of the ML estimation are not met, we expected that it would provide biased estimates of standard error and the chi-square statistic.
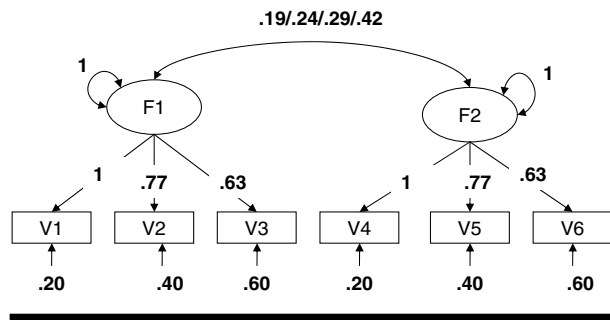
Figure 2

Phenotypic two correlated latent factor model.

Note: The diagram contains the true population values based on the simulation model. The three correlations between the latent factors correspond respectively to the conditions A40–A40, A60–A40, A60–A60 and A35–C25.

We expected the MLR estimation with correction for clustering to reduce the bias in all conditions.

The percentage relative bias was used to evaluate the accuracy of the chi-square statistic, parameter estimates, and standard errors. The percentage relative bias is computed as $((\hat{\theta}-\theta)/\theta)*100$, where $\hat{\theta}$ is the mean estimated value of chi-square, parameter estimates and standard errors across replications and θ is the true population value. Figure 2 shows the fitted model including the true parameter values (factor loadings, residual variances, and variance–covariance of the latent factors) as chosen in the simulation study. The latent factors were scaled by fixing the first factor loading to 1, and the variances of the latent factors were freely estimated. The means of the observed variables were zero in the model. The expected value of the chi-square statistic equals the number of degrees of freedom (df) asymptotically. The df of the model equal 14, that is, 27 (observed statistics) minus 13 (estimated parameters). To assess the effects of the standard errors, we compare the means of the standard error over replications with the standard deviations of the parameter estimates over replications. The mean standard deviation provides the criterion in assessing the accuracy of the standard errors, as given that the assumptions of the ML estimation are satisfied, the latter should equal the former asymptotically.

We considered both the mean chi-square, and the distribution of the chi-square by comparing the proportion of replications for which the critical values are exceeded with the expected proportions (.05 and .01) under a chi-square distribution, when fitting the correct model.

Finally, we fitted the phenotypic factor model to the data of a single member of the family (Twin 1). This enables the comparison of the results for data with different degrees of dependency with the results for data that are independent.

Results

The Monte Carlo procedure in Mplus (which both simulates and analyzes the data) showed that the true

parameter values were correctly recovered in more than 95% of the replications in all simulation conditions. For the replications with sample size equal to 2000, 4.5% of the replications of the 40–40 heritability condition, 3.7% of the 60–40 condition, 4.1% of the 60–60 condition, and 2.3% of the A35–C25 condition were excluded from further analyses, because of nonconvergence or inadmissible estimates. For replications with family size equal to 500 all replications converged and gave acceptable parameter estimates.

Table 1 shows the relative bias in the estimation of the chi-squared statistic.

The analyses of dependent family data as independent using ML result in an overestimation of the chi-square statistic. In addition, the probability of the true model being rejected (error type I) tends to be larger than expected under a central chi-squared distribution. For the first set of simulations, the bias increases with higher heritability, and it tends to be larger when twins and parents are analyzed. For the second set of simulations, with a general C component, the size of the bias in the chi-square for FS = 2 was equivalent to that of the A60–A60 condition. However, the pattern of results for FS was inverted, so that the amount of bias increased with the inclusion of parents and spouses in the sample. However, it should be noted that overall the bias in the chi-squared statistic is quite small, ranging from 2% to 11% across all conditions. The same pattern of results is shown for sample sizes 2000 and 500, although the overestimation of the chi-square for FS = 2 appears to be larger for the smaller sample size. This result does not seem to be a consequence of the sample size, but probably a random product of the simulation process. This can be inferred from the observation of the relative bias when only Twin 1 was analyzed (FS = 1). For all heritability conditions, for $N = 500$ and FS = 1 the relative bias was slightly positive, whereas for $N = 2000$ and FS = 1 the relative bias was negative. The difference in relative bias between conditions FS = 2 and FS = 1 is equivalent for $N = 2000$ and $N = 500$ (4.1/4.1, 4.9/5.2, 6.1/6.3 for the three heritability conditions respectively).

When the clustered data are analyzed using the complex estimation in Mplus, the bias due to dependency is corrected for all conditions, and the estimates of the chi-square return to the levels obtained under independent sampling (see FS = 1 in Table 1), and the distribution of the chi-square, in terms of nominal and observed error rates, is well approximated.

Table 2 shows the percentage of relative bias of the parameter estimates. In all, 13 parameters were freely estimated in the factor model (six residual variances, four factor loadings, two latent factor variances and one correlation). The tables show the mean bias across all parameter estimates in relative and absolute values. The parameter estimates were perfectly recovered across all conditions, with percentages of absolute bias always below 1%.

Table 1

Bias on Chi-Square Statistic and Its Distribution

Family size	Estimation method (Mplus)[a]	Percentage relative bias on chi-square				Observed proportion for expected proportion = .05				Observed proportion for expected proportion = .01			
		A40–A40	A60–A40	A60–A60	A35–C25	A40–A40	A60–A40	A60–A60	A35–C25	A40–A40	A60–A40	A60–A60	A35–C25
N(families) = 2000													
6: Twins, parents and spouses (*N* = 11,806)[b]	General	2.600	3.807	4.914	11.143	.067	.070	.079	.110	.022	.022	.022	.038
	Complex	−1.907	−1.829	−1.829	−2.400	.057	.054	.057	.50	.016	.015	.012	.12
4: Twins and parents (*N* = 7992)	General	4.314	5.836	7.300	9.350	.072	.078	.086	.094	.021	.022	.023	.032
	Complex	−1.757	−1.743	−1.764	−2.271	.045	.051	.052	.052	.016	.017	.016	.022
2: Twins (*N* = 4000)	General	2.221	3.000	3.950	3.936	.060	.062	.069	.068	.013	.014	.017	.016
	Complex	−1.264	−1.364	−1.321	−1.457	.044	.047	.048	.046	.009	.013	.012	.009
1: Twin 1 (*N* = 2000)	General	−1.907	−1.943	−2.143	−1.764	.044	.046	.043	.049	.015	.015	.013	.015
N(families) = 500													
2: Twins (*N* = 1000)	General	4.279	5.436	6.500		.071	.076	.079		.008	.010	.015	
	Complex	0.643	0.779	0.764		.062	.058	.057		.006	.008	.008	
1: Twin 1 (*N* = 500)	General	0.179	0.150	0.129		.047	.046	.045		.011	.011	.010	

Note: Percentage relative bias on the chi-square (14 degrees of freedom) statistic over the replications and proportion of replications for which the critical value is exceeded.
[a] General: ML. Complex: MLR with cluster correction.
[b] Actual sample size, smaller than number of families × family size due to missingness.

Irene Rebollo, Marleen H. M. de Moor, Conor V. Dolan, and Dorret I. Boomsma

Table 2

Percentage Relative Bias on Parameter Estimates

Family size	Mean bias (mean |bias|)[a]			
	A40–A40	A60–A40	A60–A60	A35–C25
N(families) = 2000				
6: Twins, parents and spouses (*N* = 11,806)[b]	−0.126(0.231)	−0.119(0.234)	−0.122(0.234)	0.021(0.077)
4: Twins and parents (*N* = 7992)	−0.129(0.253)	−0.117(0.255)	−0.119(0.256)	0.026(0.089)
2: Twins (*N* = 4000)	−0.095(0.248)	−0.089(0.268)	−0.093(0.263)	0.054(0.115)
1: Twin 1 (*N* = 2000)	−0.167(0.234)	−0.154(0.262)	−0.156(0.267)	−0.002(0.103)
N(families) = 500				
2: Twins (*N* = 1000)	−0.267(0.288)	−0.276(0.297)	−0.283(0.309)	
1: Twin 1 (*N* = 500)	−0.388(0.472)	−0.394(0.489)	−0.412(0.514)	

Note: Not affected by estimation method.

[a] Mean bias: Mean of parameter bias across all parameter estimates. Mean \|bias\|: Mean of parameter bias in absolute value across all parameter estimates.

[b] Actual sample size, smaller than number of families × family size due to missingness.

Table 3 shows the percentage of relative bias on the standard errors averaged across all parameter estimates.

When the family data are analyzed using ML the standard errors of the parameter estimates tend to be underestimated, judging by the mean relative values. For the first set of simulations, the bias is larger when the heritability is greater, and when only twins are analyzed, compared to conditions where parents and spouses are included. For the second set of simulations, the pattern is again inverted so that the bias is lowest when only twins are analyzed. The size of the bias for the A35–C25 condition for FS = 2 is similar to that for the A60–A60 condition, whereas the size of the bias gets larger for FS = 4 or 6. However, again the bias is quite small: under no condition does the average exceed 15%, or the maximum exceed 20%. The bias fluctuates across parameter estimates, as it can be appreciated in the difference between the relative and the absolute value of the mean bias. Furthermore, the pattern of results and the size of the bias are comparable for sample sizes equal to 2000 and 500.

When the complex estimation method of Mplus is used to estimate the factor model, the negative bias in the standard error is corrected across all conditions, so that the mean percentage relative bias shows values close to zero or slightly positive, that is, comparable to the values obtained when the data are independent, that is, data of single family members. However, the size of bias of the standard error was not equal across all parameter

Table 3

Percentage Relative Bias on Standard Errors Across All Parameter Estimates

Family Size	Estimation method (Mplus)[c]	Mean bias (mean |bias|)[a]			
		A40–A40	A60–A40	A60–A60	A35–C25
N(families) = 2000					
6: Twins, parents and spouses (*N* = 11,806)[b]	General	−0.496(2.011)	−0.408(2.182)	−0.689(2.581)	−2.206(4.243)
	Complex	0.125(1.996)	0.345(1.734)	0.541(1.813)	0.519(2.208)
4: Twins and parents (*N* = 7992)	General	−0.405(2.335)	−0.744(2.722)	−1.055(3.007)	−2.567(4.120)
	Complex	0.382(1.548)	0.416(1.562)	0.406(1.423)	0.155(1.508)
2: Twins (*N* = 4000)	General	−0.786(2.448)	−1.122(2.664)	−1.429(3.013)	−1.932(3.367)
	Complex	−0.086(2.172)	−0.132(2.234)	−0.172(2.167)	−0.289(2.137)
1: Twin 1 (*N* = 2000)	General	−0.581(1.748)	0.024(1.837)	−0.053(1.976)	−0.194(2.131)
N(families) = 500					
2: Twins (*N* = 1000)	General	−0.609(1.494)	−0.902(1.844)	−1.065(2.155)	
	Complex	−0.029(1.546)	0.005(1.673)	0.208(1.764)	
1: Twin 1 (*N* = 500)	General	−0.999(1.930)	−0.969(1.927)	−0.952(1.972)	

Note: [a] Mean bias: Mean bias across the standard error of all parameter estimates. Mean \|bias\|: Mean bias in absolute value across the standard error of all parameter estimates.

[b] Actual sample size, smaller than number of families × family size due to missingness.

[c] General: ML. Complex: MLR with cluster correction.

Table 4

Percentage Relative Bias on Standard Errors for Variances and Covariance of the Latent Factors

Family size	Estimation method (Mplus)[c]	Mean bias(mean |bias|)[a]				% of standard error bias in the model due to the bias on variances and covariance of the latent factors			
		A40–A40	A60–A40	A60–A60	A35–C25	A40–A40	A60–A40	A60–A60	A35–C25
		N(families) = 2000							
6: Twins, parents and spouses (*N* = 11,806)[b]	General	−2.743(2.743)	−3.526(3.526)	−5.240(5.240)	−11.452(11.452)	31.48%	37.28%	46.84%	62.28%
	Complex	0.145(2.480)	−0.031(1.813)	0.088(1.911)	0.354(2.636)	28.67%	24.12%	24.32%	27.54%
4: Twins and parents (*N* = 7992)	General	−4.107(4.107)	−5.847(5.847)	−7.517(7.517)	−12.665(12.665)	40.58%	49.57%	57.68%	70.94%
	Complex	−0.956(0.956)	−0.823(0.823)	−0.920(0.920)	−0.868(1.346)	14.25%	12.15%	14.92%	20.60%
2: Twins (*N* = 4000)	General	−3.625(4.638)	−5.093(5.093)	−7.192(7.192)	−8.912(8.912)	43.73%	44.11%	55.09%	61.08%
	Complex	−0.803(3.234)	−1.028(3.000)	−1.756(3.542)	−1.612(3.403)	34.36%	30.99%	37.71%	36.74%
1: Twin 1 (*N* = 2000)	General	−1.040(1.295)	−0.495(1.033)	−1.047(1.047)	−0.974(0.974)	17.079%	12.97%	12.23%	10.55%
		N(families) = 500							
2: Twins (*N* = 1000)	General	−1.537(1.537)	−2.729(2.729)	−4.081(4.081)		23.73%	34.14%	43.70%	
	Complex	1.171(1.559)	1.416(1.986)	1.644(2.218)		23.28%	27.38%	29.01%	
1: Twin 1 (*N* = 500)	General	1.566(1.566)	1.701(1.701)	1.834(1.834)		18.73%	20.37%	21.46%	

Note: [a] Mean bias: Mean bias across the standard error of variances and covariance of the latent factors. Mean |bias|: Mean bias in absolute value across the standard error of variances and covariance of the latent factors.

[b] Actual sample size, smaller than number of families × family size due to missingness.

[c] General: ML. Complex: MLR with cluster correction.

estimates. The estimates of the variances and covariance of the common factors display the largest bias. Table 4 shows the mean percentage relative bias and the percentage of total bias in these three parameters.

When the family clustered data are analyzed using ML estimation, the standard error of the parameter estimates of the latent factors are consistently underestimated between 2.7 and 7.5% across conditions in the first set of simulations, and between –8.9 and 11.45% in the second set of simulations. The pattern across heritability and family size conditions resembles the one observed for all the parameters. In the first set, the bias is larger for higher heritability values and for FS = 2 or 4. For the second set, the bias is smallest for FS = 2.

When the Mplus complex estimation method is used to correct for clustering, the bias is reduced to values comparable to those obtained with FS = 1, remaining under 2%.

Discussion

The results of the simulation study indicated there is a tendency for the chi-square statistic to be overestimated, and the standard errors of the parameters underestimated when standard ML estimation is used to analyze family clustered data, and dependent individuals are treated as independent cases. Furthermore, the distribution of the chi-square is affected, resulting in an increase in Type I errors. When fitting a model with common latent factors, most of the bias is localized in the standard error of the variances and covariances of the common factors.

Figure 3 shows an overview of the results across conditions. It can be observed how, under ML estima-

tion, the positive bias in the chi-square and the negative bias in standard error increase with family resemblance. Family size and the source of family resemblance interact. In the absence of C, samples with twins, parents and spouses show the lowest bias, whereas in the presence of C, samples with only twins show the lowest bias. The effect of family resemblance on the bias is independent of sample size. The results for A60–A60 and A35–C35 conditions suggest that it is the total amount of family resemblance, and not its nature, that determines the amount of bias produced in the standard error when only twins are analyzed.

Figure 3 clearly depicts how the use of the corrected MLR (complex) estimation reduces the bias across all conditions to levels observed when only independent cases are considered.

These results were concordant with the hypotheses, except for the bias on the chi-square across family size conditions in the first set of simulations. We expected the largest bias in both standard error and chi-square in the FS = 2 condition. Whereas the standard error bias follows the predicted pattern, the chi-square bias turned out to be lowest in this condition, and highest for FS = 4. However, with such a small amount of bias, we do not consider these slight differences between FS conditions to be a cause of concern.

It must be noted that the scope of the present study is somewhat limited by the assumptions of the simulated model and sampling conditions. Further simulations will reveal whether varying certain conditions will give rise to significantly different results. Such conditions may pertain to different kinds of phenotypic analysis or theoretical models, larger number of observed variables, the inclusion of other
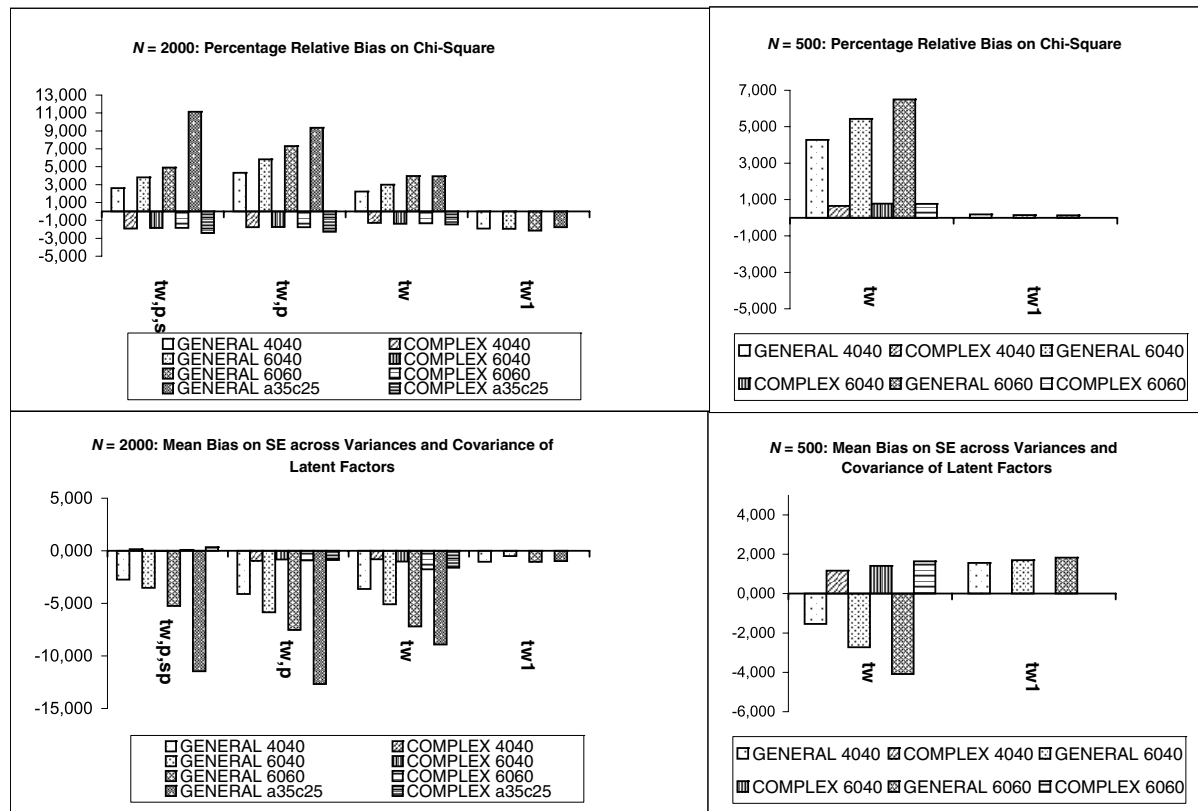
Figure 3

Summary of results.

Note: Percentage relative bias on chi-squared statistic, and mean bias on the standard error of parameter estimates of the latent factors F1 and F2 are depicted across conditions.
   tw: twins, p: parents, sp: spouses, tw1: Twin 1.

kinds of relatives (siblings, children of twins, and so on), or unbalanced number of MZ and DZ families. On the other hand, the good outcome of the MLR correction due to clustering might be due to aggregatability. According to Muthén and Satorra (1995) a two-level factor model is aggregatable if the factor loading matrices are equal on the within-cluster and between cluster levels. In the present paper we have used a common pathway model that assumes that the measurement model for the E factor and the A factor has the same structure. The results might not apply when the variance decomposition of the phenotype is better explained by an independent pathway model, for which the factor structures of A and E differ significantly.

Although the results appear to be quite acceptable, the method for correcting for dependency (i.e., the 'complex' option in Mplus) is approximate as the cluster unit used is the family. This implies that the correction is undertaken assuming that the dependency is homogeneous across family members, so the same correction is applied, for example, to MZ and DZ families, or to spouses, parents and twins. However, the problem of different model structures for different family members could to some extent be handled in a multiple group analysis. Alternative

strategies to obtain adjusted estimates are possible: for example, replication methods like bootstrap (Laplante & Hebert, 2001), or treating each family as a case and fitting the model in the diagonal within-person part of the matrix, estimating the off diagonal elements as nuisance parameters (for a detailed explanation of this method using Mx [Neale et al., 2003], see van der Sluis et al., in press). However, the complex option in Mplus does arguably provide the easiest, if approximate, way to correct for dependency.

Multilevel structural equation modeling and hierarchical linear mixed modeling is often used to handle data clustering (Hox & Maas, 2001; Muthén & Satorra, 1995). With these methods the variance-covariance matrix is decomposed into within and between cluster components. Using the family as a cluster implies that all the variance due to the differences among the members of the same family is analyzed in the within-level part of the model, whereas the variance due to the differences between different families is analyzed in the between-level part of the model. In real settings the exact sources of resemblance between family members may be unknown, and therefore it is also unknown which components of the variance (genetic, shared environment, nonshared environment, assortative mating,

etc.) are placed in the within and between parts of the model. Under these conditions, it is unclear how the parameter estimates within each level can be interpreted, or how they can be related to the true values obtained under phenotypic analysis with independent sampling. This problem is exacerbated when data from extended pedigrees are analyzed, because the sources of differences and resemblance between each pair of members of the family are different, and the family is treated as a homogeneous cluster in multilevel modeling. Therefore, multilevel modeling assumes that the family members are statistically equivalent, whereas with the multivariate approach it would be possible to give different parameter values to different family members.

Summing up, through the use of robust estimates the bias due to family dependency becomes practically negligible. Even if no correction is applied, the point estimates are correct, and if the chi-square statistic indicates a good model fit ($p > .05$), it is certain that the same conclusion would be attained with independent data. However certain precautions should be taken when interpreting the results of uncorrected solutions. It is possible that certain parameters appear to be statistically significant, whereas they may actually be nonsignificant in the analysis of truly independent observations. If the chi-square statistic indicates poor model fit ($p < .05$), there is a small probability that a good model fit would have been obtained with independent data, because of the increase in Type I error due to dependency. If family data are used for phenotypic analysis with no correction for the inherent dependency, one should take into account that the nature of the family resemblance for the given trait, and the family composition of the sample interact to result in larger or smaller amounts of bias in the standard error. When variation in the phenotype under study is exclusively due to genetic and nonshared environmental effects, an extended family sample will result in a smaller bias, whereas if family resemblance in the phenotype under study is also due to shared environment, cultural transmission or assortative mating, the use of an extended family sample will result in a larger amount of bias.

In conclusion, the present study shows that the gains of using the richness of family data from twin registers for phenotypic analysis outweighs the relatively small drawbacks of the slight bias in standard error and chi-square statistic. More importantly, it shows that irrespective of the source of family resemblance or dependency, the bias is successfully corrected by the MLR estimation with clustering correction of the Mplus program.

Endnote

1   The Mplus scripts used to generate and analyze the data are available upon request to the first author.

References

Azzelini, A. (1996). *Statistical inference based on the likelihood.* London: Chapman & Hall.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Boomsma, D. I. (1998). Twin registers in Europe: An overview. *Twin Research, 1,* 34–51.

Bouchard, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological Differences. *Journal of Neurobiology, 54,* 4–45.

Brown, A. S., Schaefer, C. A., Wyatt, R. J., Begg, M. D., Goetz, R., Bresnahan, M. A., Harkavy-Friedman, J., Gorman, J. M., Malaspina, D., & Susser, E. S. (2002). Paternal age and risk of schizophrenia in adult offspring. *American Journal of Psychiatry, 159,* 1528–1533.

Busjahn, A. (2002). Twin registers across the globe: What's out there in 2002? *Twin Research, 5,* V–VI.

Eaves, L. J., Heath, A., Martin, N. G., Maes, H. H. M., Neale, M., Kendler, K. S., Kirk, K., & Corey, L. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30000 study of twins and their relatives. *Twin Research, 2,* 62–80.

Edwards, K. L., Austin, M. A., Newman, B., Mayer, E., & Selby, J. V. (1994). Multivariate-analysis of the Insulin-Resistance Syndrome (Irs) in women. *Circulation, 89,* 934.

Eley, T. C., Bolton, D., O'Connor, T. G., Perrin, S., Smith, P., & Plomin, R. (2003). A twin study of anxiety-related behaviours in pre-school children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 44,* 945–960.

Falconer, D. S. (1989). *Introduction to quantitative genetics* (3rd ed.). Harlow: Longman.

Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8,* 157–174.

Jonnal, A. H., Gardner, C. O., Prescott, C. A., & Kendler, K. S. (2000). Obsessive and compulsive symptoms in a general population sample of female twins. *American Journal of Medical Genetics, 96,* 791–796.

Kirk, K. M., Hickie, I. B., & Martin, N. G. (1999). Fatigue as related to anxiety and depression in a community-based sample of twins aged over 50. *Social Psychiatry and Psychiatric Epidemiology, 34,* 85–90.

Laplante, B., & Hebert, B. P. (2001). An introduction to the use of linear models with correlated data. *Canadian Studies in Population, 28,* 287–311.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25,* 267–316.

Muthén, L. K., & Muthén, B. O. (2005). *Mplus user's guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling.* (6th ed.). VCU Box 900126, Richmond, VA 23298: Department of Psychiatry.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers.

Tozzi, F., Aggen, S. H., Neale, B. M., Anderson, C. B., Mazzeo, S. E., Neale, M. C., & Bulik, C. M. (2004). The structure of perfectionism: A twin study. *Behavior Genetics, 34*, 483–494.

van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (in press). Sex differences on the Dutch WAIS-IIIR. *Intelligence*.

Wade, T. D., Wilkinson, J., & Ben Tovim, D. (2003). The genetic epidemiology of body attitudes, the attitudinal component of body image in women. *Psychological Medicine, 33*, 1395–1405.