Statistical Power to Detect Genetic and Environmental Influences in the Presence of Data Missing at Random

Eske M. Derks,[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*

We study the situation in which a cheap measure (X) is observed in a large, representative twin sample, and a more expensive measure (Y) is observed in a selected subsample. The aim of this study is to investigate the optimal selection design in terms of the statistical power to detect genetic and environmental influences on the variance of Y and on the covariance of X and Y. Data were simulated for 4000 dizygotic and 2000 monozygotic twins. Missingness (87% vs. 97%) was then introduced in accordance with 7 selection designs: (i) concordant low + individual high design; (ii) extreme concordant design; (iii) extreme concordant and discordant design (EDAC); (iv) extreme discordant design; (v) individual score selection design; (vi) selection of an optimal number of MZ and DZ twins; and (vii) missing completely at random. The statistical power to detect the influence of additive and dominant genetic and shared environmental effects on the variance of Y and on the covariance between X and Y was investigated. The best selection design is the individual score selection design. The power to detect additive genetic effects is high irrespective of the percentage of missingness or selection design. The power to detect shared environmental effects is acceptable when the percentage of missingness is 87%, but is low when the percentage of missingness is 97%, except for the individual score selection design, in which the power remains acceptable. The power to detect D is low, irrespective of selection design or percentage of missingness. The individual score selection design is therefore the best design for detecting genetic and environmental influences on the variance of Y and on the covariance of X and Y. However, the EDAC design may be preferred when an additional purpose of a study is to detect quantitative trait loci effects.

Questionnaires can be a cost-effective way to obtain information on a wide variety of phenotypes (e.g., behavior, health, and social environment). Because of the relatively low costs, it is feasible to phenotype large numbers of subjects. Therefore, twin registries often include survey data collected with questionnaires in twins and their family members. For some purposes, however, it may be necessary to collect more expensive phenotypic (e.g., endophenotypic or biological) measures. For example, in gene finding studies, high costs are involved in the collection of DNA samples and the subsequent genotyping. Endophenotyping may include endocrine measures, assessment of neurocognitive measures, or assessment of brain structure volumes and functioning with Magnetic Resonance Imaging (MRI). Expensive phenotypes are sometimes collected through psychiatric interviews or 24-hour ambulatory recordings of cardiovascular functions. Because of the high costs associated with such measures, the number of subjects that can be tested is often limited. This may force one to phenotype selected subjects from those in the large representative sample. Such selection should be optimized by selecting the most informative cases given the objective of the study.

The purpose of the present article is to investigate the precision of parameter estimates in various selection designs in the context of multivariate genetic covariance structure modeling of monozygotic (MZ) and dizygotic (DZ) twin data (Martin & Eaves, 1977). Specifically, we envisage the situation in which we want to estimate the genetic and environmental covariance structures of phenotypes X and Y. Relatively cheap measures of phenotype X are available in a large representative sample of twin pairs. Phenotype Y, in contrast, is expensive to measure, and can only be measured in a selected subsample. The question is: how should we select cases from the representative samples so that we retain the greatest possible statistical power, while ensuring that the estimates are unbiased?

Sib-pair selection in quantitative trait loci (QTL) analyses (Eaves & Meyer, 1994; Gu et al., 1996;

Received 3 July, 2006; accepted 16 November, 2006.

Address for correspondence: Eske M. Derks, Vrije Universiteit, Department of Biological Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands. E-mail: em.derks@psy.vu.nl

Risch & Zhang, 1995) is designed to enrich the sample with siblings who share zero or two alleles identically by descent. The present objective is different, namely to obtain estimates of the genetic and environmental contributions to the variance of X and Y, and to the covariance between X and Y. To achieve this aim given the limitations (i.e., limited means to phenotype Y), it would seem rational to select individuals who score extremely high or low on X. Provided that the correlation between X and Y is greater than zero, subjects with extreme scores on X will also have, on average, more extreme scores on Y. The linear relationship of X on Y can be estimated well with extreme high and low values on Y, as in combination these determine the orientation of the regression line in the regression of X on Y. The loss of power to detect a correlation between X and Y, as a result of missingness, can be very small indeed in this situation (Dolan et al., 2005). The present case is more complicated, as we have twin pairs, and therefore four rather than two variables. Although it may be expected that the selection of subjects with extreme scores on X will provide the best statistical power, a number of questions can be raised with respect to the selection design. Should we select concordant and/or discordant twin pairs? Should we select on a twin pair or on an individual basis? The introduction of missingness due to selection does not pose an estimation problem, as full information (raw data) maximum likelihood (FIML) estimation can be used if the data are missing at random (MAR) or missing completely at random (MCAR), in the sense of Rubin (1976; see also Little & Rubin, 2002; Schafer & Graham, 2002). We briefly explain the concepts of MCAR and MAR. The distributions for the missingness (R) can be classified according to the nature of the relationship between the missingness and the data (Schafer & Graham, 2002). Let the complete data matrix Z_{com} be partitioned as $Z_{com} = (Z_{obs}, Z_{mis})$, where Z_{obs} and Z_{mis} are the observed and missing parts, respectively. Missing data are MAR if the distribution of missingness does not depend on Z_{mis},

$$p(R|Z_{com}) = p(R|Z_{obs}). \quad [1]$$

In other words, the probability of missingness depends only on the observed part of the data, and not on the missing part. For example, suppose that blood pressure was observed in a sample of 1000 subjects, and that some additional data were collected in all subjects who obtained a blood pressure score of one standard deviation (*SD*) above average. Here, the probability of missingness is 1, given that a subject scores below the cut-off, and the probability of missingness is 0, given that a subject scores above the cut-off. The data are MAR, because the missingness depends on observed data only. This example is relevant to the present undertaking, where we have observed X, and on the basis of X we select twin pairs for phenotyping with respect to Y.

A special case of MAR is MCAR. In the previous example, the data would be MCAR if a random sample of the 1000 subjects is invited to participate, and the probability of nonparticipation is not related to the trait of interest. In other words, the distribution of missingness does not depend on Z_{mis} or Z_{obs},

$$p(R|Z_{com}) = p(R). \quad [2]$$

When data are MAR or MCAR, a number of methods are available to deal with the missingness. These involve multivariate analysis of all available data with FIML estimation, imputation of the missing data (Little & Rubin, 2002) and data weighting (Heath et al., 1998; Little & Rubin, 2002).

In the present article, we apply FIML in seven selection designs. In five of these designs, selection gives rise to data MAR. In the sixth design, we optimize the number of MZ and DZ twins using the derivations of Visscher (2004). In the seventh design, which we include for reasons of comparison, the data are MCAR. We present the results of genetic covariance structure modeling using simulated twin data, with missingness due to selection according to the seven designs. The percentage of missingness was set at 87% and 97%. We simulated data according to two different etiological models, and fitted the models using FIML estimation in our own FORTRAN program. This program is available at http://www.tweelingenregister.org/nederlands/onderzoek/missing_at_random.htm. At this website, we also included an example of the R script that was used for the data simulation. We report the means and standard errors of the estimates of genetic and environmental parameters in each design. We also consider the power to detect additive genetic, dominant genetic, and shared environmental effects.

Methods

A Description of the Seven Selection Designs

Designs 1 to 5 are based on the selection of twins who score lowly or highly on the (cheap) phenotypic test X. The first design considered here is one that was used in a study on attention problems (AP) and attention-deficit/hyperactivity disorder (ADHD; Derks et al., 2006). Twin pairs were selected if (i) both members obtained low AP checklist scores, or (ii) at least one of the members obtained a high AP checklist score. The lower threshold was not very extreme, but only 25% of the twin pairs that obtained a low AP score was randomly selected. In selection design 2 (extremely concordant; EC), twin pairs are selected if both siblings scored extremely high, or if both siblings scored extremely low. Selection design 3 is the extreme concordant and discordant (EDAC) design. In this design, pairs are selected if the two members of a twin pair are discordant, or if both members score extremely low or high. In selection design 4, only

extremely discordant pairs (ED) are selected. In selection design 5, twin pairs are selected based on individual scores rather than on a pair-wise basis. Specifically, a complete twin pair is selected if at least one of the twins scores extremely low or high. In selection design 6, we optimize the number of MZ and DZ twin pairs according to the derivations of Visscher (2004). After establishing the optimal number of MZ and DZ pairs, subjects were randomly selected from the total population of MZ and DZ twins. Because Visscher provided derivations for the optimal MZ:DZ ratio for the ACE model only, this design was not included when data were simulated according to an ADE model. Finally, in selection design 7 (MCAR), twin pairs are selected completely at random. In addition to analyzing the data obtained in these seven selection designs, we also analyzed the complete dataset (i.e., no selection, and therefore no missingness).

Data Simulation

All data were simulated using routines in the freely available R program (Venables et al., 2002). The data were multivariate normally distributed, with unit variances and zero means. In most European countries, the number of DZ twins is about twice the number of MZ twins. Therefore, we chose a 1:2 ratio for the number of MZ:DZ twins. The number of MZ twin pairs was 2000, and the number of DZ twin pairs was 4000. The number of replications was 250. To study the effect of the amount of missingness that is introduced, the percentage of missingness was either 87% or 97%. To establish a constant percentage of missingness irrespective of the selection design, the thresholds were allowed to vary in the seven selection designs (see Tables 2a–2d). Because of the different twin covariances in MZ and DZ twins, the percentages of missingness may vary between these groups. For example, in the extreme discordant design and given that phenotype X is a heritable trait, more DZ than MZ twins will be selected.

The influence of the relative contributions of genetic and environmental factors to individual differences in ADHD can be inferred from the different

Table 1

Population Covariance Matrix (MZ/DZ Twins)

	X_1	Y_1	X_2	Y_2
X_1	1	.56/.56	.70/.45	.56/.36
Y_1	.56/.56	1	.56/.36	.70/.45
X_2	.70/.30	.56/.24	1	.56/.56
Y_2	.56/.24	.70/.30	.56/.56	1

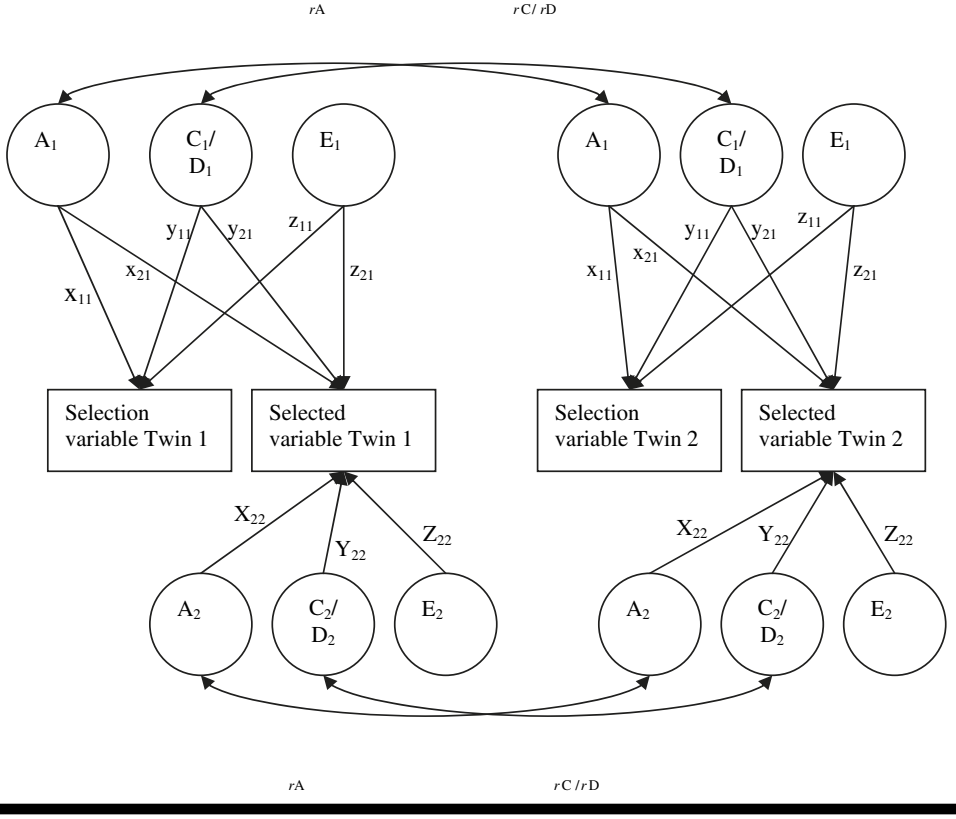

Figure 1

Bivariate path model for twin data in which the selection variable is assessed in the complete sample and the selected variable in a subsample.

Note: A = Additive genetic influences; C = Shared environmental influences; D = Dominant genetic influences; E = Nonshared environmental influences; *r*A = 1 (.5) in MZ (DZ) twins; *r*D = 1(.25) in MZ (DZ) twins; *r*C = 1 in MZ (DZ) twins.

level of genetic relatedness of MZ and DZ twins (Neale & Cardon, 1992). The variance may be due to additive genetic effects (A), dominant genetic effects (D) or shared environmental effects (C), and nonshared environmental (E) effects. The genetic effects (A and D) correlate 1 in MZ twins. In DZ twins, A correlates .5 and D correlates .25. C correlates 1 in both MZ and DZ twins. E or nonshared environmental effects are, by definition, uncorrelated. All uncorrelated measurement error, if present, is absorbed in the E term. Note that estimating C and D at the same time is not possible in a design using only data from MZ and DZ twins reared together. The decomposition of the phenotypic covariance matrix was based on Cholesky decompositions as shown in Figure 1. The variation in phenotype X (i.e., the variable on which the selection was based), phenotype Y (i.e., the variable in which missingness is introduced), and the covariance between X and Y is influenced by A, C or D, and E. Simulation was based on two etiological models: the ADE model and the ACE model. In the ADE model, variation in the phenotypes X and Y was for 50% explained by A, for 20% by D, and for 30% by E. The covariation between X and Y was for 71% explained by A and for 29% by D.

In the ACE model, A, C, and E explained 50%, 20% and 30%, respectively, of the phenotypic variance in X and Y. A and C explained 71% and 29%, respectively, of the phenotypic covariance between X and Y. In Table 1, we report the theoretical covariance matrices in MZ and DZ pairs for the ADE (below diagonal) and ACE (above diagonal) models. X_1 and Y_1 refer to variable X and Y as observed in the first-born while X_2 and Y_2 refer to variable X and Y as observed in the second-born.

Statistical Analyses

The genetic model fitting was carried out in our own FORTRAN program. In the case of the ADE model, parameter estimates were obtained by fitting a multivariate ADE model, and the statistical power to detect D on the variance of Y was obtained by fixing the factor loadings y_{21} and y_{22} at zero (see Figure 1). The power to detect influences of D on the covariance between X and Y was obtained by fixing the factor loading y_{21} at zero. Likewise, in the ACE model, parameter estimates were obtained by fitting a multivariate ACE model, and the statistical power to detect A and C on the variance of Y was obtained by fixing the factor loadings x_{21} and x_{22}, and the loadings y_{21} and y_{22}, respectively. The power to detect A and C on the covariance between X and Y was obtained by fixing the loadings x_{21} and y_{21}, respectively.

The null distribution of the likelihood ratio test for the significance of the genetic and environmental influences on the *covariance* of X and Y follows a χ^2 (1 *df*) distribution. However, because of the implicit constraints on the parameters in the Cholesky decomposition, the null distribution of the likelihood ratio test for the genetic and environmental influences on the *variance* of Y is not the expected central χ^2 (2*df*; Carey, 2004; Dominicus, Skrondal, et al., 2006). Rather it is a mixture of χ^2 distributions, differing in *df*s (see Stram & Lee, 1994). To obtain some insight into the nature of this mixture, we performed a small simulation. The results suggested strongly that the null-distribution is asymptotically a $\chi^2_{(1)}$–$\chi^2_{(2)}$ mixture with mixing proportions approximately equal to 50:50. These results tally with those of Stram and Lee (1994). The critical values associated with this mixture was estimated at 5.138 (α = .05, *df* = 2) using a R program.

Table 2a

Standardized Parameter Estimates in ADE Model (87% Missingness)

		Complete sample	Concordant low + individual high	Extreme concordant	EDAC	Extreme discordant	Individual score selection	MCAR
Threshold low		—	–.44	–.9	–1	–.4	–1.78	—
Threshold high		—	1.64	.9	1	.4	1.78	—
Mean NMZ/NDZ		2000/4000	269/526	386/433	344/459	118/679	240/551	267/532
Variance component	True value							
A variance X	.50	.50 (.05)	.50 (.05)	.50 (.05)	.50 (.05)	.50 (.05)	.50 (.05)	.50 (.05)
A covariance XY	.40	.40 (.04)	.40 (.07)	.40 (.08)	.40 (.07)	.41 (.11)	.40 (.06)	.41 (.08)
A variance Y	.50	.50 (.06)	.47 (.12)	.47 (.14)	.48 (.12)	.49 (.17)	.47 (.12)	.49 (.13)
D variance X	.20	.20 (.05)	.20 (.05)	.20 (.05)	.20 (.05)	.20 (.05)	.20 (.05)	.20 (.05)
D covariance XY	.16	.16 (.04)	.16 (.07)	.16 (.09)	.16 (.08)	.16 (.08)	.16 (.07)	.15 (.08)
D variance Y	.20	.20 (.06)	.23 (.12)	.23 (.15)	.22 (.13)	.21 (.14)	.23 (.12)	.21 (.14)
E variance X	.30	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)
E covariance XY	.00	.00 (.01)	.00 (.02)	.00 (.02)	.00 (.02)	.00 (.01)	.00 (.01)	.00 (.02)
E variance Y	.30	.30 (.01)	.30 (.02)	.30 (.02)	.30 (.02)	.30 (.03)	.29 (.03)	.30 (.02)

Note: A = additive genetic effects, D = dominant genetic effects; E = nonshared environmental effects; EDAC = extreme discordant and concordant; MCAR = missing completely at random.

Table 2b
Standardized Parameter Estimates in ACE Model (87% Missingness)

		Complete sample	Concordant low + individual high	Extreme concordant	EDAC	Extreme discordant	Individual score selection	Optimal NMZ:NDZ	MCAR
Threshold low		—	–.45	–1	–1.02	–.3	–1.77	—	—
Threshold high		—	1.65	1	1.02	.3	1.77	—	—
Mean NMZ/NDZ		2000/4000	264/524	334/463	323/481	164/640	249/555	312/488	266/531
Variance component	True value								
A variance X	.50	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)
A covariance XY	.40	.40 (.02)	.40 (.04)	.40 (.06)	.40 (.05)	.40 (.04)	.40 (.04)	.40 (.05)	.40 (.05)
A variance Y	.50	.50 (.03)	.49 (.08)	.49 (.08)	.50 (.08)	.49 (.07)	.49 (.07)	.49 (.08)	.49 (.08)
C variance X	.20	.20 (.03)	.20 (.03)	.20 (.03)	.20 (.03)	.20 (.03)	.20 (.03)	.20 (.03)	.20 (.03)
C covariance XY	.16	.16 (.02)	.16 (.04)	.16 (.05)	.16 (.04)	.15 (.06)	.16 (.03)	.16 (.04)	.16 (.04)
C variance Y	.20	.20 (.02)	.21 (.06)	.21 (.07)	.21 (.06)	.20 (.08)	.21 (.06)	.21 (.07)	.21 (.07)
E variance X	.30	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)
E covariance XY	.00	.00 (.01)	.00 (.02)	.00 (.02)	.00 (.02)	.00 (.01)	.00 (.01)	.00 (.02)	.00 (.02)
E variance Y	.30	.30 (.01)	.30 (.03)	.30 (.02)	.30 (.02)	.30 (.03)	.30 (.03)	.30 (.02)	.30 (.03)

Note: A = additive genetic effects, C = shared environmental effects; E = nonshared environmental effects; EDAC = extreme discordant and concordant; MCAR = missing completely at random.

Table 2c
Standardized Parameter Estimates in ADE Model (97% Missingness)

		Complete sample	Concordant low + individual high	Extreme concordant	EDAC	Extreme discordant	Individual score selection	MCAR
Threshold low		—	–1	–1.515	–1.52	–.77	–2.375	—
Threshold high		—	2.35	1.515	1.52	.77	2.375	—
Mean NMZ/NDZ		2000/4000	75/117	109/86	107/92	11/197	62/136	65/132
Variance component	True value							
A variance X	.50	.50 (.06)	.50 (.06)	.50 (.06)	.50 (.06)	.50 (.06)	.50 (.06)	.50 (.06)
A covariance XY	.40	.40 (.04)	.41 (.09)	.41 (.12)	.41 (.11)	.39 (.16)	.41 (.08)	.41 (.12)
A variance Y	.50	.50 (.06)	.45 (.16)	.45 (.19)	.46 (.18)	.46 (.25)	.46 (.15)	.46 (.19)
D variance X	.20	.20 (.06)	.20 (.06)	.20 (.06)	.20 (.06)	.20 (.06)	.20 (.06)	.20 (.06)
D covariance XY	.16	.16 (.04)	.15 (.10)	.15 (.12)	.15 (.12)	.17 (.12)	.16 (.08)	.15 (.11)
D variance Y	.20	.20 (.05)	.24 (.17)	.24 (.20)	.24 (.19)	.28 (.22)	.24 (.15)	.24 (.18)
E variance X	.30	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)
E covariance XY	.00	.00 (.01)	.00 (.03)	.00 (.04)	.00 (.04)	.00 (.03)	.00 (.02)	.00 (.03)
E variance Y	.30	.30 (.01)	.30 (.04)	.30 (.04)	.30 (.04)	.26 (.09)	.30 (.05)	.30 (.05)

Note: A = additive genetic effects, D = dominant genetic effects; E = nonshared environmental effects; EDAC = extreme discordant and concordant; MCAR = missing completely at random.

Results

As expected, given the FIML theory, the estimates of the genetic and environmental influences on phenotypes X, Y, and on the covariance between X and Y closely resemble the simulated values in the true ADE and ACE model (Tables 2a–2d). The *SE*s of the standardized influences on phenotype Y are greater than the *SE*s of the standardized influences on phenotype X, as is to be expected, because missingness was limited to Y. The lowest *SE*s are found in the individual score selection design.

The statistical power to detect the influences of A, C, and D on the variance of Y is reported in Tables 3a and 3b for 87% and 97% missingness, respectively. Tables 3c and 3d report the statistical power to detect the influences of A, C, and D on the covariance between X and Y for the respective percentages of missingness. The decrease in statistical power as a result of the introduction of missingness can be derived from a comparison of the mean χ^2 of the selected samples with the mean χ^2 of the complete sample (i.e., the value when no missingness is introduced). A lower mean χ^2 is

Table 2d

Standardized Parameter Estimates in ACE Model (97% Missingness)

		Complete sample	Concordant low + individual high	Extreme concordant	EDAC	Extreme discordant	Individual score selection	Optimal NMZ:NDZ	MCAR
Threshold low		—	–1	–1.57	–1.57	–.67	–2.375	—	—
Threshold high		—	2.35	1.57	1.57	.67	2.375	—	—
Mean NMZ/NDZ		2000/4000	74/129	95/106	95/108	21/179	61/134	78/122	65/132
Variance component	True value								
A variance X	.50	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)
A covariance XY	.40	.40 (.02)	.41 (.07)	.40 (.08)	.40 (.08)	.40 (.06)	.41 (.06)	.41 (.09)	.41 (.09)
A variance Y	.50	.50 (.03)	.49 (.13)	.47 (.15)	.47 (.15)	.49 (.14)	.49 (.12)	.47 (.15)	.49 (.14)
C variance X	.20	.20 (.03)	.20 (.02)	.20 (.02)	.20 (.02)	.20 (.02)	.20 (.02)	.20 (.02)	.20 (.03)
C covariance XY	.16	.16 (.02)	.16 (.05)	.16 (.07)	.16 (.07)	.15 (.10)	.16 (.05)	.16 (.08)	.15 (.07)
C variance Y	.20	.20 (.03)	.21 (.10)	.22 (.12)	.22 (.12)	.21 (.15)	.21 (.10)	.22 (.13)	.21 (.12)
E variance X	.30	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)
E covariance XY	.00	.00 (.01)	.00 (.03)	.00 (.04)	.00 (.04)	.00 (.03)	.00 (.02)	.00 (.03)	.00 (.03)
E variance Y	.30	.30 (.01)	.30 (.05)	.31 (.04)	.31 (.04)	.30 (.08)	.30 (.05)	.31 (.05)	.30 (.05)

Note: A = additive genetic effects, C = shared environmental effects; E = nonshared environmental effects; EDAC = extreme discordant and concordant; MCAR = missing completely at random.

associated with a lower statistical power. In addition to the mean χ^2, we included the proportion of simulations in which the null hypothesis (i.e., no effect of A, C, or D) was rejected.

The statistical power to detect D on the variance of Y and on the covariance of X and Y is substantially decreased as a result of the selection in all seven designs. For example, the power to detect a dominant genetic component which explains 20% of the variance of Y equals .99 in the absence of missingness. After the introduction of 87% missingness, the statistical power drops to .27–.68 (see Table 3a), with the individual score selection design showing the highest power. Likewise, the statistical power to detect D on the covariance of X and Y with 87% missingness drops from .96 to .43–.69. As for the detection of D on the variance of Y, the highest power is obtained with the individual score selection design. Generally, the statistical power to detect C is higher than the statistical power to detect D. However, the pattern of results is similar, with the highest power being obtained in the individual score selection design. The power to detect A approaches 1, irrespective of the percentage of missingness or selection design. Clearly the effect size of A is too large to pick up any differences in power.

Discussion

In the current article we focused on the situation in which a relatively cheap measure (X) is measured in a

Table 3a

Effect of Missingness on the Statistical Power to Detect Additive Genetic Effects (A), Dominant Genetic Effects (D), and Shared Environmental Effects (C) on the Variance of Y (87% Missingness)

	Detection of D (in ADE model)		Detection of C (in ACE model)		Detection of A (in ACE model)	
	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power
Complete sample	18.14 (8.35)	.99	77.37 (16.18)	> .99	430.13 (36.17)	> .99
Concordant low + individual high	6.67 (4.69)	.56	22.85 (9.75)	.99	101.68 (20.87)	> .99
Extreme concordant	3.78 (3.07)	.30	12.79 (6.26)	.87	77.91 (17.05)	> .99
Extreme concordant and discordant	5.78 (4.84)	.46	17.83 (8.04)	.96	83.01 (17.95)	> .99
Extreme discordant	3.80 (3.19)	.27	9.02 (5.53)	.74	110.97 (19.21)	> .99
Individual score selection	8.52 (5.81)	.68	29.76 (10.71)	> .99	122.09 (21.04)	> .99
Optimal NMZ:NDZ	—	—	14.99 (7.20)	.97	87.26 (17.23)	> .99
MCAR	4.33 (3.72)	.30	15.53 (7.37)	.94	79.86 (16.69)	> .99

Note: Statistical power is power with alpha = .05, *df* = 2; MCAR = missing completely at random.

Table 3b

Effect of Missingness on the Statistical Power to Detect Additive Genetic Effects (A), Dominant Genetic Effects (D), and Shared Environmental Effects (C) on the Variance of Y (97% Missingness)

	Detection of D (in ADE model)		Detection of C (in ACE model)		Detection of A (in ACE model)	
	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power
Complete sample	17.74 (7.55)	.99	76.67 (17.16)	> .99	434.37 (35.32)	> .99
Concordant low + individual high	2.78 (2.43)	.16	8.61 (5.07)	.75	39.80 (11.12)	> .99
Extreme concordant	1.84 (1.89)	.06	4.97 (3.63)	.41	31.53 (10.29)	> .99
Extreme concordant and discordant	2.23 (2.24)	.09	5.66 (4.23)	.48	31.81 (10.49)	> .99
Extreme discordant	2.25 (2.12)	.09	2.81 (2.48)	.14	33.55 (10.25)	> .99
Individual score selection	3.68 (3.16)	.27	11.48 (6.24)	.86	47.71 (12.71)	> .99
Optimal NMZ:NDZ	—	—	5.32 (4.05)	.44	28.74 (10.55)	> .99
MCAR	1.92 (1.95)	.08	5.18 (4.00)	.43	26.44 (10.00)	> .99

Note: Statistical power is power with alpha = .05, *df* = 2; MCAR = missing completely at random.

Table 3c

Effect of Missingness on the Statistical Power to Detect Additive Genetic Effects (A), Dominant Genetic Effects (D), and Shared Environmental Effects (C) on the Covariance Between X and Y (87% Missingness)

	Detection of D (in ADE model)		Detection of C (in ACE model)		Detection of A (in ACE model)	
	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power
Complete sample	14.00 (7.19)	.96	61.86 (15.39)	> .99	379.52 (34.94)	> .99
Concordant low + individual high	5.49 (4.31)	.58	20.96 (9.44)	> .99	101.65 (20.88)	> .99
Extreme concordant	3.05 (2.70)	.32	11.51 (6.11)	.88	77.72 (17.05)	> .99
Extreme concordant and discordant	4.78 (4.49)	.47	16.30 (7.81)	.98	82.83 (17.95)	> .99
Extreme discordant	3.44 (3.08)	.35	7.58 (5.47)	.71	110.91 (19.26)	> .99
Individual score selection	7.14 (5.25)	.69	27.62 (10.49)	> .99	122.07 (21.06)	> .99
Optimal NMZ:NDZ	—	—	13.48 (6.81)	.96	87.08 (17.19)	> .99
MCAR	3.47 (3.23)	.34	13.86 (7.14)	.94	79.75 (16.68)	> .99

Note: Statistical power is power with alpha = .05, *df* = 1; MCAR = missing completely at random.

Table 3d

Effect of Missingness on the Statistical Power to Detect Additive Genetic Effects (A), Dominant Genetic Effects (D), and Shared Environmental Effects (C) on the Covariance between X and Y (97% Missingness)

	Detection of D (in ADE model)		Detection of C (in ACE model)		Detection of A (in ACE model)	
	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power	Mean χ^2 (*SD*)	Statistical power
Complete sample	13.85 (6.87)	.96	61.75 (15.16)	> .99	382.91 (32.85)	> .99
Concordant low + individual high	2.38 (2.26)	.24	8.19 (5.06)	.80	39.78 (11.13)	> .99
Extreme concordant	1.60 (1.81)	.10	4.51 (3.48)	.48	31.52 (10.29)	> .99
Extreme concordant and discordant	1.95 (2.13)	.16	5.21 (4.06)	.56	31.80 (10.48)	> .99
Extreme discordant	2.02 (2.05)	.17	2.18 (2.17)	.18	33.54 (10.26)	> .99
Individual score selection	3.22 (3.00)	.34	10.98 (6.16)	.92	47.70 (12.72)	> .99
Optimal NMZ:NDZ	—	—	4.77 (3.89)	.54	28.63 (10.60)	> .99
MCAR	1.51 (1.66)	.09	4.69 (3.87)	.50	26.35 (10.01)	> .99

Note: Statistical power is power with alpha = .05, *df* = 1; MCAR = missing completely at random.

large representative sample, and a more expensive measure (Y) is measured in a subset of the sample. Little and Rubin (2002) have shown that, provided that the missing data are either MAR or MCAR, FIML produces an unbiased estimate of the population covariance matrix. The goal of this article was to investigate the best selection design in terms of the statistical power to detect genetic and environmental influences on the variance of Y and on the covariance of X and Y.

Standard errors of the estimates of genetic and environmental influences were smallest in the individual score selection design. The better performance of this design was also reflected in the results of the power analyses, as it resulted in the highest statistical power to detect genetic and environmental influences on the variance of Y and on the covariance between X and Y. Clearly, for the purpose of covariance decomposition, informative twin pairs should not be selected on a pair-wise basis but rather on an individual basis. What could be the explanation for this finding? In the individual score selection design, twins with a moderate score are included if their co-twin has an extreme score. It is likely that the inclusion of moderate scores improves the estimation of the regression of X on Y. To test this hypothesis, we simulated data and estimated the variances and covariances in MZ and DZ twins. Eighty-seven per cent missingness was introduced by including a random sample of 5% of the subjects while the remaining 8% of the twin pairs were selected when they were extreme concordant. This design is a combination of the MCAR and the EC design. With this design, the standard errors of the variances and covariances were lower compared to both the EC design and the MCAR design. The explanation for this finding is that the inclusion of subjects with both moderate and extreme scores increases the precision of the estimations.

Although the individual score selection design was the best design in this study, most studies are not limited to the decomposition of phenotypic variance into latent genetic and environmental influences. An additional purpose may be the detection of effects of measured genetic polymorphisms. For this purpose, the ED and EDAC designs are more suitable (Eaves & Meyer, 1994; Gu et al., 1996; Risch & Zhang, 1995). With these designs, additive genetic effects, which usually explain more variation than shared environmental or dominant genetic effects do not pose a problem. The power to detect shared environmental factors is especially problematic in the ED design, but is acceptable in the EDAC design, as long as the percentage of missingness is not too extreme. The power to detect D is not only very poor in the ED and EDAC designs, but also in the individual score selection design. Therefore, the EDAC design may be a good alternative, if one is interested both in traditional genetic covariance structure analysis and in QTL analyses.

The fact that the statistical power to detect A and C is satisfactory, even with a percentage of missingness of 87% (resulting in a sample size of 800 pairs), is promising for longitudinal studies, in which attrition may lead to missing phenotypic data in a subset of the sample. Usually, nonresponse rates will be much lower than 87%, which suggests that the power to detect genetic and environmental influences is good, given that the original (total) sample size is not very small. Considering the effect of missingness as a result of attrition, it should be kept in mind that we restricted ourselves to the discussion of response models under the assumption that the data are MAR or MCAR. When the data are actually missing not at random (MNAR), maximum likelihood estimation cannot correct for the missingness, and genetic and environmental influences may either be overestimated or underestimated, depending on the nonresponse model (Dominicus, Palmgren, et al., 2006; Taylor, 2004).

The results of the present study should be interpreted in the light of the following limitations. First, we assumed that the data are either MAR or MCAR. This does not affect the results in the present studies, because the selection strategies that were applied guaranteed MAR. In other situations, however, the mechanism giving rise to missingness may be unknown which would result in data MNAR. As discussed above, in these instances, parameters may be biased. Second, we assumed that phenotypes X and Y are multivariate normally distributed. If the true distribution of X is not normal, for example, as a result of censoring or truncation, the selection may not be optimal and the decrease in statistical power will be greater than observed in the current analyses.

Acknowledgments

This work was supported by NWO Grant numbers 575-25-006, 575-25-012, and 904-57-94 (Boomsma, P.I.), NIMH Grant number MH58799 (Hudziak, P.I.), and by the Centre for Neurogenomics and Cognition Research (CNCR) of the Vrije Universiteit, Amsterdam.

References

Carey, G. (2004). Cholesky problems. *Behavior Genetics, 34,* 633.

Derks, E. M., Hudziak, J. J., Dolan, C. V., Ferdinand, R. F., & Boomsma, D. I. (2006). The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Comprehensive Psychiatry, 47,* 116–122.

Dolan, C., van der Sluis, S., & Grasman, R. (2005). A note on normal theory power calculation in SEM with data missing completely at random. *Structural Equation Modeling-A Multidisciplinary Journal, 12,* 245–262.

Dominicus, A., Palmgren, J., & Pedersen, N. L. (2006). Bias in variance components due to nonresponse in

twin studies. *Twin Research and Human Genetics, 9,* 185–193.

Dominicus, A., Skrondal, A., Gjessing, H. K., Pedersen, N. L., & Palmgren, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics, 36,* 331–340.

Eaves, L., & Meyer, J. (1994). Locating human quantitative trait loci - Guidelines for the selection of sibling pairs for genotyping. *Behavior Genetics, 24,* 443–455.

Gu, C., Todorov, A., & Rao, D. C. (1996). Combining extremely concordant sibpairs with extremely discordant sibpairs provides a cost effective way to linkage analysis of quantitative trait loci. *Genetic Epidemiology, 13,* 513–533.

Heath, A. C., Madden, P. A. F., & Martin, N. G. (1998). Assessing the effects of cooperation bias and attrition in behavioral genetic research using data-weighting. *Behavior Genetics, 28,* 415–427.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley and Sons.

Martin, N. G., & Eaves, L. J. (1977). Genetic analysis of covariance structure. *Heredity, 38,* 79–95.

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

Risch, N., & Zhang, H. P. (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science, 268,* 1584–1589.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–590.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics, 50,* 1171–1177.

Taylor, A. (2004). The consequences of selective participation on behavioral-genetic findings: Evidence from simulated and real data. *Twin Research, 7,* 485–504.

Venables, W. N., Smith, D. M., & R Development Core Team. (2002). *An introduction to R*. Bristol, United Kingdom: Network Theory Limited.

Visscher, P. M. (2004). Power of the classical twin design revisited. *Twin Research, 7,* 505–512.