Genetic Association in Multivariate Phenotypic Data: Power in Five Models

Camelia C. Minica,[1] Dorret I. Boomsma,[2] Sophie van der Sluis,[3] and Conor V. Dolan[1]

[1] *Department of Psychology, FMG, University of Amsterdam, The Netherlands.*

[2] *Biological Psychology, VU University, The Netherlands.*

[3] *Functional Genomics Section, Department of Clinical Genetics, Center for Neurogenomics and Cognitive Research (CNCR), VU University and VU University Medical Center, The Netherlands*

This article concerns the power of various data analytic strategies to detect the effect of a single genetic variant (GV) in multivariate data. We simulated exactly fitting monozygotic and dizygotic phenotypic data according to single and two common factor models, and simplex models. We calculated the power to detect the GV in twin 1 data in an ANOVA of phenotypic sum scores, in a MANOVA, and in exploratory factor analysis (EFA), in which the common factors are regressed on the genetic variant. We also report power in the full twin model, and power of the single phenotype ANOVA. The results indicate that (1) if the GV affects all phenotypes, the sum score ANOVA and the EFA are most powerful, while the MANOVA is less powerful. Increasing phenotypic correlations further decreases the power of the MANOVA; and (2) if the GV affects only a subset of the phenotypes, the EFA or the MANOVA are most powerful, while sum score ANOVA is less powerful. In this case, an increase in phenotypic correlations may enhance the power of MANOVA and EFA. If the effect of the GV is modeled directly on the phenotypes in the EFA, the power of the EFA is approximately equal to the power of the MANOVA.

Keywords: power, association, multivariate, univariate, exact data simulation, genetic covariance structures.

Well-established twin registries, such as the Scandinavian twin registers (Peltonen, 2003), Netherlands Twin Register (Boomsma et al., 2006), the UK Adult Twin Register (Spector & Williams, 2006), and the Brisbane Adolescent twin study (Wright & Martin, 2004), contain a wealth of multivariate phenotypic data, relating to many different phenotypes, and often observed at multiple occasions. Developments in genotyping technology have resulted in the addition of measured genetic information to these databases (Boomsma et al., 2002; Willemsen et al., 2010). The availability of genetic data has allowed researchers to shift their focus from family-based genetic covariance structure modeling (Martin & Eaves, 1977; Neale & Cardon, 1992) to the detection of individual gene effects in linkage and/or association studies (e.g., Balding, 2006; Fulker et al., 1999; Hirshhorn & Daly, 2005; Olson et al., 1999; Kettunen et al., 2009; Perola et al., 2007; Vink & Boomsma, 2002; Yang et al., 2010). Given the presence of multivariate phenotypic data, the question arises under which conditions a multivariate analysis is preferable to univariate analyses in studying the role of a given genetic variant (GV).

In linkage analyses, multivariate modeling was considered both for substantive reasons and for statistical power advantages that multivariate data conferred (e.g., Allison et al., 1998; Amos et al., 2001; Boomsma, 1996; Boomsma & Dolan, 1998; Boomsma & Dolan, 2000; Evans & Duffy, 2004; Hottenga & Boomsma, 2008; Martin et al., 1997). To date, population-based association studies have focused mainly on the relationship between a measured measured GV and a univariate phenotype. In the case of psychological phenotypes, this phenotype is often a sum score (i.e., the sum calculated across all items of a phenotypic instrument), or a case-control affection status dichotomy. In genetic association studies, however, the power advantages of multivariate data are also of interest, especially as the contributions of individual genetic variants to the phenotypic variance are commonly assumed to be small (Evans, 2008; Gordon & Finch, 2005). To date, three studies have addressed the question of the power to detect GVs using multivariate data. In this paper, we briefly discuss these studies, and we contribute to this area by examining the power to detect a GV in genetic covariance structures based on the single and two common factor models and models for repeated measures.

Ferreira and Purcell (2009) considered the power of a multivariate test (MANOVA) based on Wilk's Lambda given varying number of phenotypes (5, 10,

Received 15 September, 2010; accepted 01 October, 2010.

Address for correspondence: Conor V. Dolan, Department of Psychology, FMG, University of Amsterdam, Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands. E-mail: c.v.dolan@uva.nl

and 20), of which a varying number were affected by the GV. They also varied the positive intercorrelations between the phenotypes. They found that the multivariate test was more powerful than univariate tests, with (1) increasing correlations among the phenotypes and (2) increasing number of phenotypes affected (i.e., by the GV) increasing the power. However, they noted a sharp loss of power of the multivariate test when all phenotypes were affected by the GV. This loss in power is exacerbated by increasing phenotypic correlations. Their results are consistent with previous results obtained in linkage analysis (Allison et al., 1998; Evans & Duffy, 2004; Ferreira, 2006), and with the statistical literature on MANOVA (Cole et al., 1994).

Medland and Neale (2010) considered the single factor models with 3 or 5 indicators, in unrelated cases and in sib pairs (Fulker et al., 1999). They varied the effect of the GV in the factor model such that it was (1) part of the common factor, thus conveying its effect via the factor loadings on all variables; or (2) common to all phenotypes, but not conveyed via the factor; or (3) present only in a single phenotype, or in some (but not all) phenotypes; or (4) it was present in some phenotypes, but with opposite effects. Medland and Neale (2010) studied the power to detect the GV in the factor model, in which the GV affected all phenotypes via the factor (one degree of freedom (DF) test), or directly affected all phenotypes (a DF = 3 or DF = 5 test). They also considered the power conferred by the univariate tests based on simple sum scores and factor scores (Lawley & Maxwell, 1971). They varied other important aspects such as the magnitude of the factor loadings and the degree of missingness. Based on their figures 1a and 1b (Medland & Neale, 2010, p. 237), the main conclusion is that their combined multivariate approach (where the GV effect is conveyed via the common factor, or the GV affects the phenotypes directly) was almost universally as powerful as, or, depending on specific circumstances, more powerful than, the univariate tests using sum scores or factor scores.

Van der Sluis et al. (in press) discussed the power to detect the effects of GVs in uni- and multidimensional common factor models. They contrasted the power in these models to the sum score model, in the situation that the sum score is not a sufficient statistic (i.e., the univariate sum score entails a loss of information relative to the multivariate data). They showed that the use of the sum score generally entails a loss of power, except in specific circumstances. In addition, they discussed how violations of measurement invariance across multiple samples, or with respect to the GV itself, affect the power to detect GVs. Violations of measurement invariance with respect to the GV itself (i.e., direct effects of the GV on one or more phenotypes in the model, instead of GV effects that are common to all phenotypes and mediated by (genetic) common factors) resulted in notable loss of power in the sum score model and incorrectly specified factor models.

The present aim is to contribute to this work on the power to detect genetic association in multivariate data. We discuss five models that one may encounter in family-based genetic covariance structure modeling of MZ and DZ twin data (Neale & Cardon, 1992): genetic factor models with single or multiple genetic factors underlying the covariance among a set of phenotypes, and two variations on the simplex models, which have been used to analyze repeated measures (Eaves et al., 1986; Boomsma & Molenaar, 1987). In each model, the effect of the GV is specified as part of an additive genetic factor, so that its effect on the phenotypes is mediated by the additive genetic factor. We consider situations in which the GV affects all phenotypes, and situations in which the effect is limited to a subset of the phenotypes. The single common factor model has been considered previously in the studies by Medland and Neale (2010) and van der Sluis et al. (in press). The power to detect a GV in the other four models has not been considered so far.

We simulated data according to a full multivariate twin model in the five scenarios. We established the power to detect the GV in this true model, and we studied the power in four statistical models using only the data of the first twin members, that is, in genetically uninformative samples. In the following sections, we describe the five study designs and the simulation procedures in more detail. Next, we present the results, and we end the paper with a discussion.

Procedure

To calculate the power to detect the GV effect, we generated conditionally multivariate normal (i.e., conditional on the GV) MZ and DZ twin data according to the five models of interest. Next, we computed the power to detect the GV effect in the full MZ and DZ twin data, and in three statistical models in which we used only the twin 1 data (i.e., the phenotypic data and the measured GV): a univariate ANOVA in which the sum of the phenotypic measures was regressed on the GV, MANOVA in which all phenotypes were regressed on the GV, and exploratory factor analysis (EFA), in which the common factors were regressed on the GV.

We simulated multivariate data according to a multivariate ACE twin model, in which A, C, and E represent the additive genetic structure, shared, and specific environmental influences, respectively. The additive genetic structure included one or more additive genetic factors. To one of these, we added a single diallelic codominant GV (minor allele frequency of .2), and defined its effect of .25% of the variance of a given phenotype, which loaded directly on the genetic factor. Depending on the chosen additive genetic factor structure, the GV did (directly or indirectly), or did not exert an influence on any other phenotype.

The first model that we considered included a single additive genetic factor. The single factor model was considered implicitly by Ferreira and Purcell

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

(2009),[1] and explicitly by Medland and Neale (2010) and van der Sluis et al. (in press). In the present study, the GV was specified as a source of variation in the genetic factor, and so this factor mediated the relationship between the GV and the phenotypes (see Figure 1). We include it because the single factor model, as specified below, is an ideal, and because the comparison of the MANOVA and the EFA has yet to be made. The second and third models included two correlated additive genetic factors. In the second model, the GV was part of the first genetic factor, but exerted no influence on the second factor or on its indicators. These indicators are thus uninformative with respect to the effect of the GV. In the third model, the second factor was regressed on the first genetic factor. This implied that the GV of the first factor did exert an influence on the second factor, and thus on its indicators. This model may represent a latent phenotype-endophenotype relationship, in which the effect of the GV on the phenotype is mediated by the endophenotype (De Geus & Boomsma, 2001; De Geus et al., 2001). Finally, we considered two hybrid simplex-factor models for repeated measures. These models have been applied mainly in genetic covariance structure modeling of twin data (Neale & Cardon, 1992; for a linkage application, see Eaves et al., 1995 and Birley et al., 2005). We considered an ACE model with the additive genetic and environmental autoregressions, and a common shared environmental factor, and a stationary AE simplex model. In the latter, the common shared environmental factor is omitted, and background influences of A and E are stable over time. In the former, shared environmental effects decline, and the genetic effects increase. We considered 4 repeated measures, and calculated the power to detect the GV effect given that it entered the model at occasions 1, 2, 3, or 4. We consider this to be of interest, as genetic innovation variance is often attributed to the action of new genetic effects (Eaves et al., 1986; Gillespie et al., 2004), which may include the effects of measured GVs. The simplex-factor model is similar to twin models established in analysis of IQ data in children, with a decreasing role of shared environment and increasing genetic influences (e.g., Hoekstra et al., 2007). The stationary AE simplex model is consistent with results one would expect in twin studies of IQ conducted in young adults. Further details on the sim-

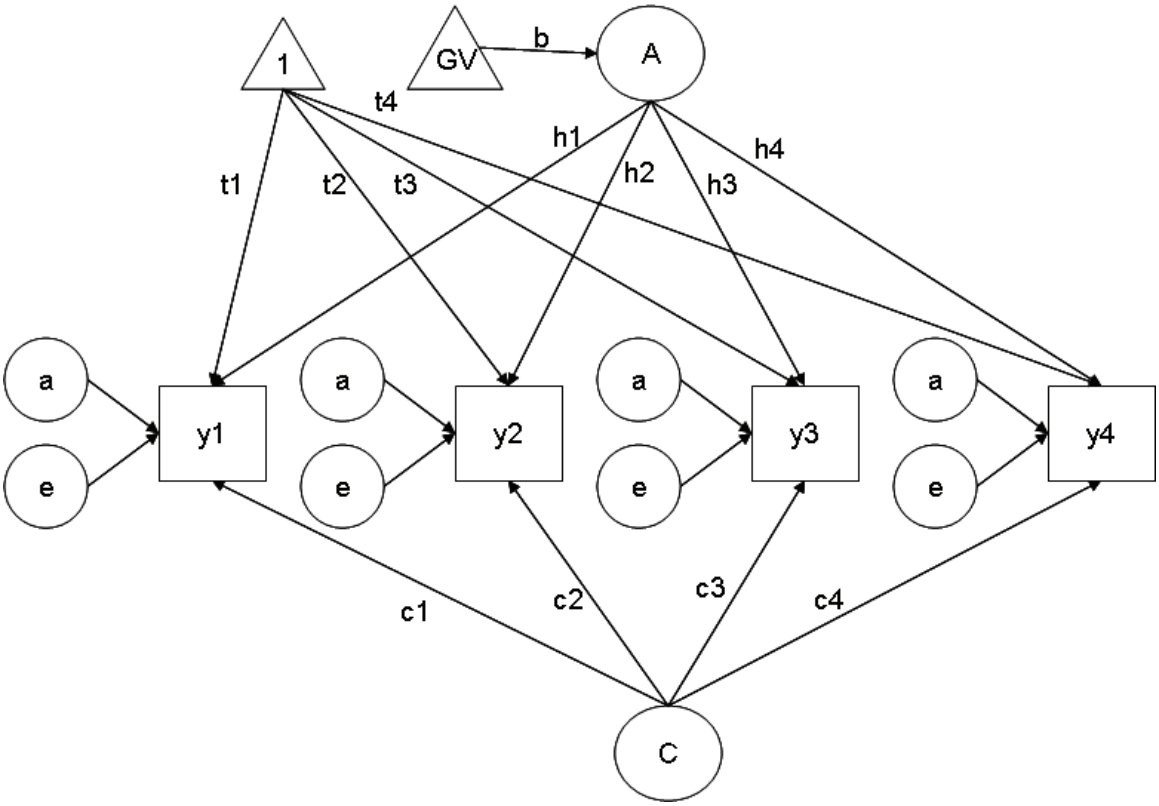

Figure 1

Path diagram for the common factor model with 4 phenotypes. The triangles represent fixed regressors (i.e., the GV and the unit vector). The parameters t1 to t4 are intercepts, the parameter b is the effect of the GV on the common genetic factor. The GV enters the model via common genetic factor A and affects the indicators y_1 to y_4. The effect size of the GV was defined as .25% of the variance of y1.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

ulation settings are given in the tables and path diagrams below.

Given these models, we varied (1) the number of phenotypic measures, and (2) the parameter values that accounted for genetic and environmental contributions to phenotypic variance. The parameter values are supposed to be typical of results one may obtain in genetic covariance structure modeling. We provide these details below. Throughout we used exact data simulation (van de Sluis et al., 2007).[2] We simulated the data using MVRNORM in R (R-core development team, 2005),[3] under the assumptions that mating is random, and the GV is in Hardy-Weinberg equilibrium. Given the diallelic GV, the total MZ and DZ sample sizes were distributed over 3 MZ groups (three pairs of identical genotypes) and 9 DZ groups (3 genotypes × 3 genotypes). The distribution of the total sample size over these groups depends on the minor allele frequency, which we set to equal .2 in all studies.

We first computed the power to detect the GV effect in (A) the full multivariate twin model. We calculated the power both in the model specified correctly with respect to the role of the GV (i.e., a 1 DF test), and in the model in which all present common genetic factors were regressed on the GV, that is, an omnibus test, with DF equaling the number of genetic factors in the model. We added the power of the omnibus test because in practice one will not know the exact locus of the GV, and therefore will resort to the omnibus test. As mentioned we did not consider the possibility that the GV affects a single phenotype (Medland & Neale, 2010, did consider this possibility), we therefore limited our omnibus test to the common genetic factors. In the twin 1 phenotypic and GV data, we calculated the power in: (B) a univariate ANOVA, in which each univariate phenotype was regressed on the GV; (C) a univariate ANOVA in which the sum of the phenotypic measures was regressed on the GV; (D) MANOVA in which all phenotypes were regressed on the GV; and (E) an exploratory factor analysis (EFA), in which the phenotypic common factors were regressed on the GV. We fitted standard MANOVAs, subject to homogeneity of the conditional (i.e., on the GV) covariance matrices. We did not constrain these in the light of our information concerning the covariance structure. In specifying the EFAs, we did exploit this information to the extent that the specified dimensionality of the exploratory factor solution is consistent with the true model. We did not fit the exploratory factor model to the repeated measures data, as the autoregressive covariance structures are not compatible with an exploratory factor model (e.g., Mandys et al., 1994). As we considered only additive genetic effects, we included the GV as a covariate (rather than as a between-subject factor) in the analyses. Analyses A and E were done in Mx (Neale et al., 2003), analyses B to D were done in R. We report the power of the tests of models B to E for N = 3000, and the power of the full twin model for NMZ = 1500 and NDZ = 1500, all given an α level of .01. In the case of the single phenotype ANOVA, we also report the power for the Bonferroni corrected alpha (i.e., .01 divided by the number of phenotypes). This correction is conservative, but the differences in power between the single phenotype test and the other tests are such that the choice of correction is unlikely to have any bearing on the conclusions. We note that a resample procedure such as permutation testing is unsuited as the data simulation is exact. The alpha of .01 is unrealistic given multiple testing. However, here we were interest solely in the differences between the tests in power, not in the absolute values. However, we report the non-centrality parameters (NCPs), so that the power of the tests of association can be computed for other total sample sizes and other α levels, if the reader so desires. R scripts that can be used to this end are provided in Appendix A. We report the power in the full twin model, as our simulation and testing procedure produces this result. However, our main interest is in the sum score ANOVAs, MANOVAs, and EFAs. The comparison of the power in the full twin model with the power of the other tests is complicated by (1) the difference in number of individuals (a twin comprises two individuals), and (2) the differences in the expense of ascertainment (ascertaining twin pairs is usually more expensive than ascertainment of unrelated indi-

Table 1

Variance Components in the Four Scenarios That Were Used to Generate the Data

Phenotypic correlations	Nr. phenotypes	Scenario	Common A	Specific a^i	Common C	Specific e^i
.5	4/8	S1	.5	.1	0	.4
.2	4/8	S2	.2	.1	0	.7
.7	4/8	S3	.5	.1	.2	.2
.4	4/8	S4	.2	.1	.2	.5

Note: The total variance of each phenotype, conditional on the GV, equaled one. We provide only 4 parameter values in each scenario, as we did not vary these parameter values over the phenotypes. For instance in scenario S3, conditional on the GV, 4 (or 8) tests loaded on the common A factor with loadings equal to $\sqrt{.5}$, the genetic residual is .1. The loadings on the common shared environmental factor equaled $\sqrt{.2}$. The unshared environmental residuals equaled 0.2. So in scenario S3, the decomposition of phenotypic variance conditional on GV is $h^2 = .6$, $c^2 = .2$, and $e^2 = .2$.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Table 2

The Power, Non-Centrality Parameter, and Degrees of Freedom (in Parentheses) of Univariate and Multivariate Tests of Association Given $\alpha = .01$ In Study 1

Scenario	Nr. phenotypes	True model	ANOVA sum scores	ANOVA single phenotype	MANOVA	EFA
N		2 x 1500	3000	3000	3000	3000
S1	4	.95	.81	.56, *.39*	.59	.81
		18.04	12.02	7.51	12.01	12.00
		(1)	(1,2998)	(1,2998)	(4,2995)	(1)
	8	.97	.85	.56, *.31*	.51	.85
		19.97	13.35	7.51	13.32	13.33
		(1)	(1,2998)	(1,2998)	(8,2991)	(1)
S2	4	.99	.96	.56, *.39*	.85	.96
		29.47	18.78	7.51	18.76	18.73
		(1)	(1,2998)	(1,2998)	(4,2995)	(1)
	8	.99	.99	.56, .31	.89	.99
		38.09	25.04	7.51	24.98	24.95
		(1)	(1,2998)	(1,2998)	(8,2991)	(1)
S3	4	.91	.70	.56, *.39*	.46	.70
		15.55	9.69	7.51	9.68	9.68
		(1)	(1,2998)	(1,2998)	(4,2995)	(1)
	8	.93	*.73*	.56, *.31*	.35	*.73*
		16.56	10.18	7.51	10.16	10.17
		(1)	(1,2998)	(1,2998)	(8,2991)	(1)
S4	4	.98	.86	.56, *.39*	.67	.86
		22.11	13.66	7.51	13.64	13.63
		(1)	(1,2998)	(1,2998)	(4,2995)	(1)
	8	.99	.91	.56, *.31*	.62	.91
		26.75	15.81	7.51	15.78	15.78
		(1)	(1,2998)	(1,2998)	(8,2991)	(1)

Note: In the case of the single phenotype ANOVA, power is reported for $\alpha = .01$ and $\alpha = .01/4$ (.0025; 4 phenotypes) or $\alpha = .01/8$ (.00125; 8 phenotypes). The power for the corrected alpha is displayed in italics.

viduals). In the subsequent sections, we present the five studies in detail.

Study 1: Single Common Genetic Factor

The objective of the first simulation study is to examine the power to detect a GV that affects all phenotypes via a common polygenic factor. Specifically, we examined how the sources of phenotypic correlations and the number of measured phenotypes affect the power to detect the GV effect. In this study we supposed that a single common genetic factor or, a common genetic factor and a shared environmental factor, account for the phenotypic correlations.

We simulated MZ and DZ phenotypic data that generate precisely the means and variances predicted by the common factor model shown in Figure 1. We specified either 4 (as depicted) or 8 phenotypes loading on the additive polygenic factor (A) and a shared environmental factor (C). Additional parameters are the phenotype-specific genetic (a_i) and unique environmental (e_i) factors. We added the GV to the common genetic factor (A), which thus affected all phenotypes (y_i). The GV accounted for .25% of the variance in the first phenotype (y_1). The chosen parameter values are given in Table 1. As we did not vary the parameters over the phenotypes, the effect size of .25% also holds with respect to the other phenotypes. In fitting the models we did not constrain any parameter to be equal, even though they were.

We simulated twin data, given eight scenarios in which we varied the role of the common factor A and C, and the specific environmental effect, as shown in Table 1. The heritability of the phenotypes ranged from $h^2 = .3$ (S1 and S3) to $h^2 = .6$ (S2 and S4). The influence of the common C was absent in scenarios S1 and S2, and present in scenarios S3 and S4 ($c^2 = .2$). As shown in Table 1, the implied correlations among the phenotypes were .5, .2, .7, and .4 in scenario S1, S2, S3 and S4, respectively. Table 2 contains the results.

Table 2 shows that in the single common factor model, the ANOVA of sum scores has the same power as the exploratory factor model. Due to the equality (over the phenotypes) of factor loadings and residual variances, the factor scores and the sum

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Table 3

Variance Components, Conditional on GV, Used to Simulate Data in Study 2 (Correlated Genetic Factors) and Study 3 (Regression Of Genetic Factor 2 on Genetic Factor 1)

Correlation coefficient	Phenot. cor. within sets, between sets*	Nr. ind.	Scenario	Common A1, A2	Specific a_i	Common C	Specific e_i
ρ_{A1A2} = .77	.8, .7	3/5	S11	.3	.1	.5	.1
	.4, .33	3/5	S12	.3	.1	.1	.5
	.3, .23	3/5	S13	.3	.1	0	.6
ρ_{A1A2} = .47	.8, .64	3/5	S21	.3	.1	.5	.1
	.4, .24	3/5	S22	.3	.1	.1	.5
	.3, .14	3/5	S23	.3	.1	0	.6
ρ_{A1A2} = .25	.8, 57	3/5	S31	.3	.1	.5	.1
	.4, .17	3/5	S32	.3	.1	.1	.5
	.3, .07	3/5	S33	.3	.1	0	.6

Note: The within (between) set phenotypic correlation is among phenotypes that load on the same (different) genetic factor (factors). For instance in 3 indicator S22 scenario, the phenotypes y1 to y3 (y4 to y6) loaded $\sqrt{.3}$ on the common A1 (A2) factor, each phenotype loading $\sqrt{.1}$ on the common C factor. The residual variance of each phenotype equaled .6 (.5 due to specific environment; .1 due to specific genes). So in scenario S22, the decomposition of phenotypic variance conditional on GV is h^2 = .4, c^2 = .1, and e^2 = .5.

* Within set correlation is among phenotypes that load on the same genetic factor (y1-y2), the between set correlation is among phenotypes that load on the different genetic factors (y1-y6).

scores are perfectly correlated. Note that in Table 2, the information with respect to the single phenotype ANOVA is redundant because the number of phenotypes simulated is irrelevant in the analysis of a single phenotype. We included the power of the single phenotype ANOVA to ease comparison, and because the power associated with the Bonferroni corrected varies as a function of the number of tests (4 vs. 8).

The single phenotype ANOVA with Bonferroni corrected alpha consistently has lowest power. The NCPs of the sum score ANOVA, the MANOVA and the EFA are comparable, and thus affected similarly by the differences in parameter configuration. The lower power of the MANOVA compared to the EFA stems from the differences in DF of the associated tests. Increasing the number of indicators resulted in a consistent increase in power of the sum score ANOVA and EFA, as is to be expected as the increase in affected phenotypes in the one-dimensional model increases the GV signal. However, the increase in the number of phenotypes resulted in a decrease in power of the MANOVA in three cases, and slight increase in only scenario S2 (power .85 vs. .89). Overall the power of the MANOVA, sum score ANOVA, and EFA decreases with increasing phenotypic correlation (e.g., compare S1 and S3). Increasing the correlations increases the variance of the sum scores, and given the constant effect size, lowers the power of the test. Cole et al. (1994) explained the role of the magnitude of phenotypic correlations on the power in the MANOVA given consistent effects on the dependent variables. Specifically they showed in two dimensions that the overlap between the 95% ellipsoids increases with increasing correlation (see Cole et al., 1994, Figure 1). This results in a loss of power (see also Ferreira & Purcell, 2009).

In conclusion, in this study, the methods of choice are the EFA or the sum score ANOVA. The power of these methods is equal because the factor loadings in the EFA are equal. We refer to Medland and Neale (2010) and van der Sluis, et al. (in press) for results obtained in the same model, but with unequal loadings. The MANOVA fares relatively poorly because all phenotypes are affected by the GV and the phenotypic correlations are positive and relatively high (notably in scenario S3). The power of the EFA is relatively good because the test involves a single parameter, i.e., the latent mean difference between the genotypes on the common factor. The effect of the GV on the actual phenotypes is thus mediated by the common factor. Medland and Neale (2010) also considered the EFA in which the GV has a direct effect on the phenotype. In that case, the NCP would be the same as shown in Table 2 for the EFA, but the DFs would equal 4 or 8 (i.e., the number of phenotypes). The power of this EFA based test would then equal that of the MANOVA.

The power of the full twin model was consistently high (>.90), but the NCPs display good variation (ranging from 15.55 to 38.09). For instance, retaining the sample sizes of 2 × 1500, but changing the alpha from 1E-2 to 1E-7, reduced the power to a low of .083 (S3, 4 phenotypes) and to a high of .801 (S2, 8 phenotypes).

Study 2: Correlated Genetic Common Factors

In the second study, the model included two correlated genetic factors, of which only the first is affected by the GV. We examined how the sources of phenotypic correlations, that is, the genetic correlation and the shared environmental factor, affect the power to detect the GV effect. In addition, we explored the impact of

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

the number of phenotypic indicators (3 vs. 5 per factor) on the power. Figure 2 depicts the three indicator model. The covariances among the phenotypes are caused by two genetic correlated factors (A1 and A2) and a shared environmental factor (C). Additional parameters in the model are genetic specifics (a_i) and unshared environmental effects (e_i). The GV enters the model via the latent genetic factor A1, and so affects the indicators of A1, but affects neither A2, nor its indicators. The parameter values used to generate data for this study are given in Table 3. The GV explained .25% of the variance of the phenotype y_1. Given the parameter values, the GV explained the same amount of variance in the other indicators for the first genetic factor, but no variance in the indicators of the second genetic factor.

As we manipulated the correlations between the genetic factors (3 settings), the number of phenotypes per genetic factor (2 settings), and the parameter values (3 settings), we simulated data according to 18 scenarios. As above, we computed the power to detect the GV in the true multivariate twin model, in the univariate phenotype ANOVA, in the sum score ANOVA, in the MANOVA, and in the two factor oblique EFA. In fitting the EFA in Mx, we identified the model by fixing the loading of the first phenotypic variable on the second factor, and the last phenotypic variable on the first factor to zero. All other loadings were estimated. The common factors were standardized and allowed to correlate. For the path diagram, see Figure 3. Other identifying constraints, but these constraints in the EFA does not affect the power of the omnibus test, in which all phenotypic common factors are regressed on the GV (McDonald, 1999; Dolan et al., 2009). In the EFA, we regressed both common factors on the GV (a 2 DF test), that is, we did not exploit our knowledge of the locus of the GV in the model. We varied the number of indicators, the size of the genetic correlations, and the contribution of the common C factor. Table 4 contains the results.

The power of the sum score ANOVA is low, as expected because the phenotypes that are unaffected by the GV add only noise to the sum score. The single (affected) phenotype ANOVA is more powerful (i.e., power based the corrected alpha) than the sum score ANOVA when the phenotype intercorrelations were relatively large (e.g., S11). The NCPs of the MANOVA and the EFA are comparable, and affected similarly by the variation in parameters. However, as in Study 1, the EFA has greater power due to the difference in DF of the associated tests. In comparison with Study 1, the MANOVA fares relatively well, because the GV does not affect all the phenotypes (as in Study 1; see also Ferreira & Purcell, 2010). Note

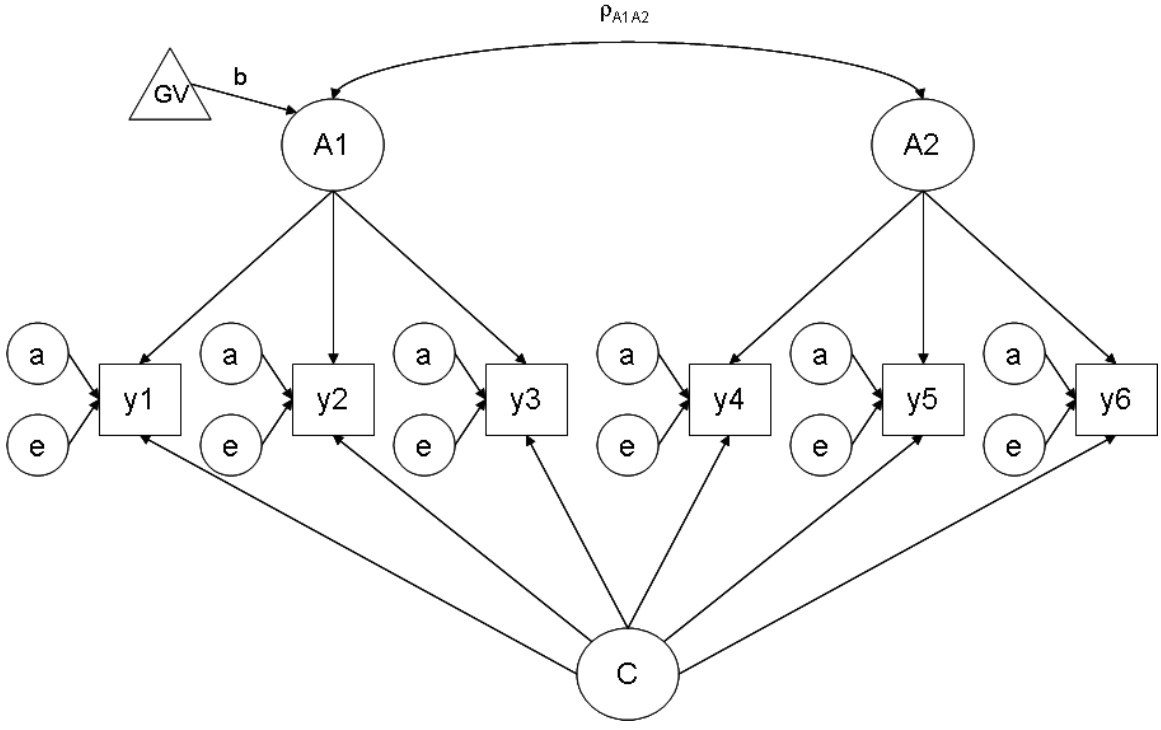

Figure 2

Path diagram of the oblique two common factor model (three indicator model). The triangle represents the GV as a fixed regressor. The unit vector, which is included to estimate intercepts is not included to avoid clutter (see Figure 1). The parameter b represents the effect of the GV. Note that the GV contributes to the variance of the first latent genetic factor A1 and affects its indicators (y1–y3), but does not affect the second common factor A2, or its indicators (y_4–y_6). The value of the correlation between A1 and A2 was varied. Parameters are not shown to avoid clutter. The effect size of the GV was defined as .25% of the variance of y1.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Table 4

Power, Non-Centrality Parameter, and Degrees of Freedom (in Parentheses) of Univariate and Multivariate Tests of Association Given $\alpha = .01$ in Study 2

ρ_{A1A2}	Scenario	n. of indicators	True model	ANOVA sum scores	ANOVA single phenotype	MANOVA	EFA
N			2×1500	3000	3000	3000	3000
$\rho_{A1A2} = .77$	S11	3	>.99,>.99 47.29 (1),(2)	.14 2.35 (1,2998)	.56, *.34* 7.51 (1,2998)	.97 30.23 (6,2993) (2)	>.99 30.15
		5	>.99,>.99 57.40 (1),(2)	.15 2.38 (1,2998)	.56, .29 7.51 (1,2998)	.98 37.14 (10,2989)	>.99 37.02 (2)
	S12	3	>.99,>.99 29.57 (1),(2)	.28 4.03 (1,2998)	.56, *.34* 7.51 (1,2998)	.77 18.02 (6,2993)	.91 18.01 (2)
		5	>.99,>.99 39.06 (1),(2)	.31 4.40 (1,2998)	.56, *.29* 7.51 (1,2998)	.85 24.35 (10,2989)	.97 24.32 (2)
	S13	3	>.99,>.99 28.48 (1),(2)	.35 4.90 (1,2998)	.56, *.34* 7.51 (1,2998)	.74 17.35 (6,2993)	.89 17.34 (2)
		5	>.99,>.99 37.83 (1),(2)	.41 5.58 (1,2998)	.56, *.29* 7.51 (1,2998)	0.83 23.60 (10,2989)	0.97 23.58 (2)
	S21	3	>.99,>.99 29.98 (1),(2)	.16 2.50 (1,2998)	.56, *.34* 7.51 (1,2998)	.79 18.62 (6,2993)	.92 18.61 (2)
		5	>.99,>.99 34.45 (1),(2)	.16 2.52 (1,2998)	.56, *.29* 7.51 (1,2998)	.78 21.55 (10,2989)	0.95 21.54 (2)
$\rho_{A1A2} = .47$	S22	3	.99,.97 23.25 (1),(2)	.32 4.50 (1,2998)	.56, *.34* 7.51 (1,2998)	.64 14.74 (6,2993)	.82 14.74 (2)
		5	>.99,>.99 28.07 (1)	.36 4.98 (1,2998)	.56, *.29* 7.51 (1,2998)	.66 18.07 (10,2989)	.91 18.07 (2)
	S23	3	.99,.97 23.61 (1),(2)	.41 5.62 (1,2998)	.56, *.34* 7.51 (1,2998)	.65 15.01 (6,2993)	.83 15.01 (2)
		5	>.99,>.99 28.86 (1),(2)	.49 6.54 (1,2998)	.56, *.29* 7.51 (1,2998)	0.69 18.77 (10,2989)	0.92 18.77 (2)
	S31	3	.99,.98 25.72 (1),(2)	.16 2.60 (1,2998)	.56, *.34* 7.51 (1,2998)	.68 15.56 (6,2993)	.85 15.55 (2)
		5	>.99,>.99 28 (1),(2)	.17 2.64 (1,2998)	.56, *.29* 7.51 (1,2998)	.61 16.91 (10,2989)	.88 16.92 (2)
$\rho_{A1A2} = .25$	S32	3	.98,.96 21.42 (1),(2)	.35 4.83 (1,2998)	.56, *.34* 7.51 (1,2998)	.59 13.70 (6,2993)	.79 13.7 (2)
		5	>.99,.98 25.25 (1),(2)	0.39 5.38 (1,2998)	.56, *.29* 7.51 (1,2998)	.59 16.31 (10,2989)	.87 16.31 (2)
	S33	3	.98,.96 22.37 (1),(2)	.46 6.15 (1,2998)	.56, *.34* 7.51 (1,2998)	.62 14.36 (6,2993)	.81 14.36 (2)
		5	>.99,.99 26.79 (1),(2)	0.54 7.26 (1,2998)	.56, *.29* 7.51 (1,2998)	.64 17.57 (10,2989)	.9 17.58 (2)

Note: The power in the true model is included for the likelihood ratio test of the correctly specified GV (1 df) and for the omnibus test, in which the 2 genetic factors are regressed on the GV (2 df). In the case of the single phenotype ANOVA, power is reported for $\alpha = .01$ and $\alpha = .01/6$ (.0016; 6 phenotypes) and $\alpha = .01/10$ (.001; 10 phenotypes). The power for the corrected alpha is displayed in italics.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

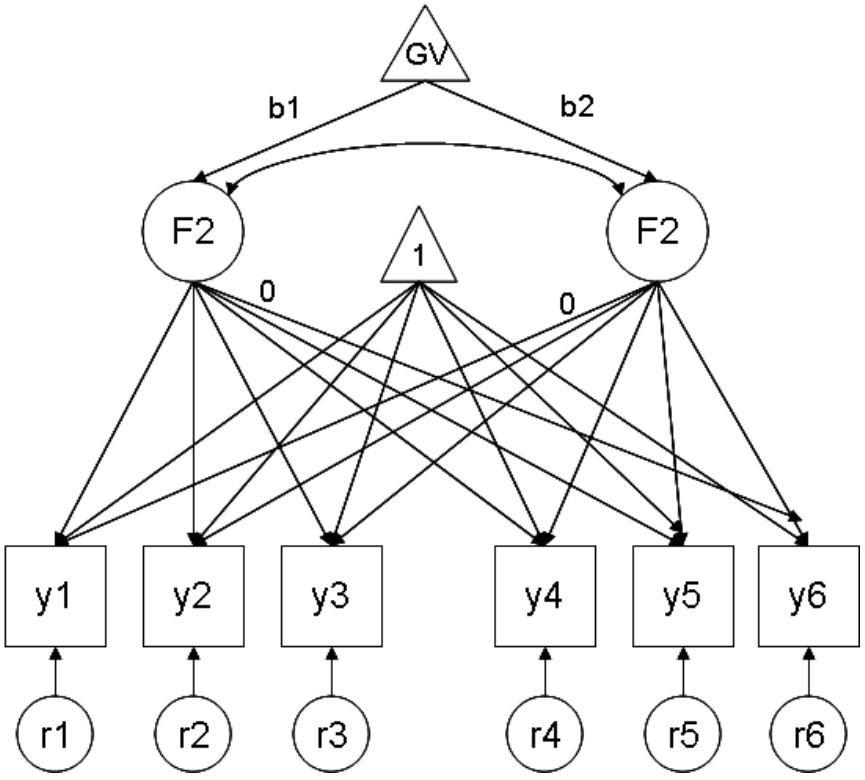

Figure 3

Exploratory (oblique) two common factor model as used in studies 2 and 3. Two factor loadings are fixed to zero (as depicted) to achieve rotational determinacy. The common factors are denoted F1 and F2, r1 to r6 represent the residuals. The triangles represent fixed regressors. The regression on the unit vector serves to estimate the intercepts, the regression on the GV estimates the effect of the GV (i.e., the parameters b1 and b2). Other parameters are not shown to avoid clutter.

that in this case (in contrast to Study 1), the increase in the phenotypic correlations resulted in an increase in power (compare S11 and S13, or S11 and S31). The presence of phenotypes not affected by the GV has a beneficial effect in MANOVA, especially when the correlations among the phenotypes are relatively high (see also Ferreira & Purcell, 2009). Cole et al. (1994) explained the role of the magnitude of phenotypic correlation on the power in the context of MANOVA, when some, but not all, dependent variables are affected (see Cole et al., 1994, Figure 3). In general, power of all tests improved by increasing the number of phenotypic indicators (from 3 to 5). In conclusion, in this study, the methods of choice are the EFA or the MANOVA. The power of the EFA is relatively good because the test involves just two parameters, i.e., two latent mean differences. As in Study 1, the effect of the GV on the phenotypes is mediated by the 2 common factors. Estimating the effect of the GV directly on the phenotypes in the EFA (a 6 or 10 DF test) would render the power of the EFA equal that of the MANOVA.

NMZ = 1500 and NDZ = 1500 afforded high power in full twin model. But again the NCPs are quite variable. Changing the alpha from 1E-2 to 1E-7 reduced the power of the 1 DF test to a low of 0.24 (scenario S32, 2 × 3 phenotypes) and to a high of >.99 (scenario S11, 2 × 5 phenotypes).

Study 3: Latent Regression Model

In the third study, we specified a latent regression model with an independent (A1) and dependent common genetic factor (A2). The GV is introduced into A1, and exerts its influence both on the indicators of A1 (i.e., via A1), and on A2, and its indicators of A2. We included a common environmental factor, and varied the details of both the shared and unshared genetic and environmental effects. As in Study 2, we considered both 3 and 5 indicators models. We simulated phenotypic data according to the model, as shown in Figure 4 (i.e., the three indicator model).

We chose parameter values such that the resulting correlations between the factors A1 and A2 equal the correlations of Study 2. The other parameters in the model are additive genetic specifics (a_i) and unique environmental effects (e_i), which contribute to phenotypic variance. The GV effect is defined with respect to the first phenotype y_1, but the GV explained the same amount of variance (.25%) in the other indicators of the first common genetic factor. Given ρ_{A1A2} = .77, the GV accounted for about 0.15% of the

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

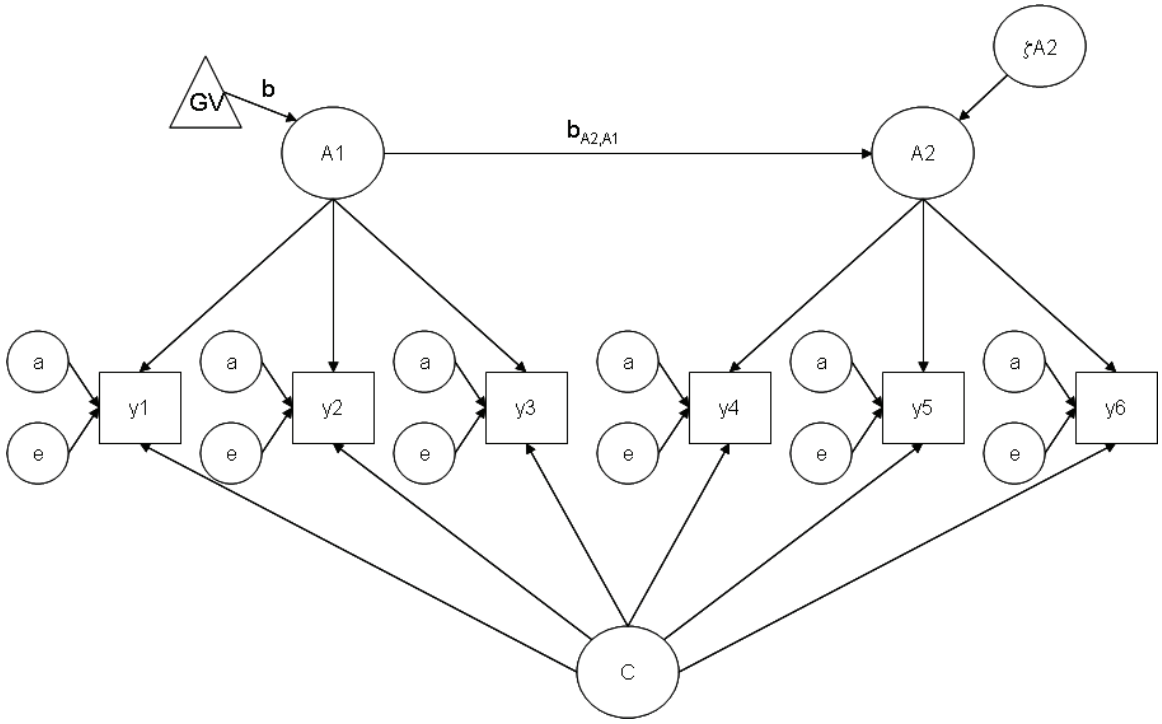

Figure 4

Path diagram for the latent genetic regression model (3 indicator model). The triangle represents the GV as fixed regressor. The unit vector, which is included to estimate intercepts is not included to avoid clutter (see Figure 1). The parameter b represents the effect of the GV. Note that the GV contributes to the variance of the first latent genetic factor A1 and affects its indicators (y_1-y_3). The GV contributes to A2 via the regression coefficient b_{A2A1}, and so also affects the indicators y_4-y_6. The value of the parameter b_{A2A1} was varied. Parameters are not shown to avoid clutter. The effect size of the GV was defined as .25% of the variance of y1.

variance in the indicators of A2, the dependent genetic factor. The parameter values used in Study 3 equaled those of Study 2, and are shown in Table 3. As we manipulated the regressions between the genetic factors (3 settings), the number of phenotypes per genetic factor (2 settings), and the parameter values (3 settings), we simulated data according to 18 scenarios. Table 5 contains the results.

The present study resembles Study 1 in that the effect of the GV is general. However, here the GV effect varied (e.g., .25% vs. .15% in S11), as did the intercorrelations among the phenotypes (see Table 3). Compared to Study 2, the sum score fares well, especially when the phenotypic intercorrelations are relatively low, and the regression relationship of A2 and A1 is relatively strong: in scenarios S11, S12, and S13, the sum score ANOVA has the greatest power. However, given a weaker regression relationship the power of the EFA is greater than the power of the sum score ANOVA. The power of the MANOVA depends on (1) the differences in the GV effect on the phenotypes (general large effects in S11, S12, and S13 in contrast to S31, S32, and S33), and the intercorrelations among the tests (generally low in S13, S23, and S33; generally high in S11, S21, and S31) (see, Cole et al., 1994). The greatest power is observed in S13 (6 phenotypes), that is, a general effect, but low phenotypic intercorrelations (.70). The lowest power is in S11 (10 phenotypes), that is, general effects and high phenotypic correlations (.27). In this scenario, the single phenotype ANOVA happens to be more powerful (.36). The NCP of the MANOVA equals that of the EFA, so it is again the difference in DF that determine the difference in power. Conducting the EFA with GV effect directly on the phenotypes (rather than being mediated by the common factors) would render the power of the EFA equal to that of the MANOVA.

The power of the 1 DF test in the full twin model is high (>.96). Changing the alpha from 1E-2 to 1E-7, reduced the power to a low of .15 (S11, 3 indicators) and to a high of .65 (S13, 5 indicators).

Study 4: Hybrid Simplex (A,E) — Factor (C) Model

In the fourth study, we considered a hybrid factor-simplex model for four occasions. We varied the occasion (t) at which the GV entered the model as part of the genetic factor (A(t)). In this model, which is shown in Figure 5, the phenotype y(t) was regressed on a latent genetic factor A(t), environmental influences common to all phenotypes C(t), and specific environmental influences E(t): y(t) = A(t) + C(t) + E(t). The stability of the phenotypic individual differences depended on the common shared environmental

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Table 5

Power, Non-Centrality Parameter, and Degrees of Freedom (in Parentheses) of Univariate and Multivariate Tests of Association Given $\alpha = .01$ in Study 3

ρ_{A1A2}	Scenario	n. of indicators	True model	ANOVA sum scores	ANOVA single phenotype	MANOVA	EFA
N			2×1500	3000	3000	3000	3000
$\beta_{A1A2} = 0.77$	S11	3	.95,.91	.55	.56 & .32, *.42 & .21*	.33	.54
			18.29	7.39	7.51 & 4.50	8.8	8.81
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	.96,.93	.56	.56 & .32, *.36 & .17*	.27	.57
			19.47	7.51	7.51 & 4.50	9.27	9.29
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
	S12	3	.97,.95	.83	.56 & .32, *.42 & .21*	.58	.78
			20.97	12.68	7.51 & 4.50	13.37	13.37
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	.98,.97	.87	.56 & .32, *.36 & .17*	.53	.83
			23.46	13.87	7.51 & 4.50	14.85	14.86
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
	S13	3	.99,.98	.91	.56 & .32, *.42 & .21*	.7	.86
			24.49	15.44	7.51 & 4.50	16.05	16.05
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	>.99,.99	.94	.56 & .32, *.36 & .17*	.67	.91
			27.86	17.58	7.51 & 4.50	18.46	18.46
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
$\beta_{A1A2} = 0.47$	S21	3	.96,.93	.38	.56 & .09, *.42 & .04*	.41	.62
			19.55	5.24	7.51 &1.50	10.16	10.17
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	.98,.95	.39	.56 & .09, *.36 & .03*	.34	.66
			21.09	5.33	7.51 &1.50	10.88	10.9
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
	S22	3	.96,.93	.68	.56 & .09, *.42 & .04*	.54	.74
			19.73	9.43	7.51 & 1.50	12.55	12.55
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	.98,.96	.74	.56 & .09, *.36 & .03*	.51	.81
			22.7	10.43	7.51 & 1.50	14.4	14.41
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
	S23	3	.98,.96	.8	.56 & .09, *.42 & .04*	.64	.82
			22.57	11.78	7.51 & 1.50	14.63	14.63
			(1) ,(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	>.99,.98	.86	.56 & .09, *.36 & .03*	.63	.89
			26.41	13.69	7.51 & 1.50	17.4	17.41
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
	S31	3	.97,.95	.29	.56 & .03, *.42 & .013*	.47	.68
			20.66	4.11	7.51 & .49	11.23	11.24
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	.98,.96	.29	.56 & .03, *.36 & .008*	.39	.72
			22.29	4.19	7.51 & .49	12.02	12.03
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
$\beta_{A1A2} = 0.25$	S32	3	.96,.93	.57	.56 & .03, *.42 & .013*	.53	.74
			19.64	7.63	7.51 & .49	12.52	12.52
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	.98,.97	.63	.56 & .03, *.36 & .008*	.51	.82
			22.76	8.51	7.51 & .49	14.52	14.52
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)
	S33	3	.99,.96	.7	.56 & .03, *.42 & .013*	.62	.81
			22.05	9.71	7.51 & .49	14.24	14.24
			(1),(2)	(1,2998)	(1,2998)	(6,2993)	(2)
		5	>.99,.98	.79	.56 & .03, *.36 & .008*	.62	.89
			26.04	11.46	7.51 & .49	17.15	17.15
			(1),(2)	(1,2998)	(1,2998)	(10,2989)	(2)

Note: The power in the true model is included for the likelihood ratio test of the correctly specified GV (1 df) and for the omnibus test, in which the 2 genetic factors are regressed on the GV (2 df). In the case of the single phenotype ANOVA, power is reported for $\alpha = .01$ and $\alpha = .01/6$ (.0016; 6 phenotypes) and $\alpha = .01/10$ (.001; 10 phenotypes). The power for the corrected alpha is displayed in italics.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

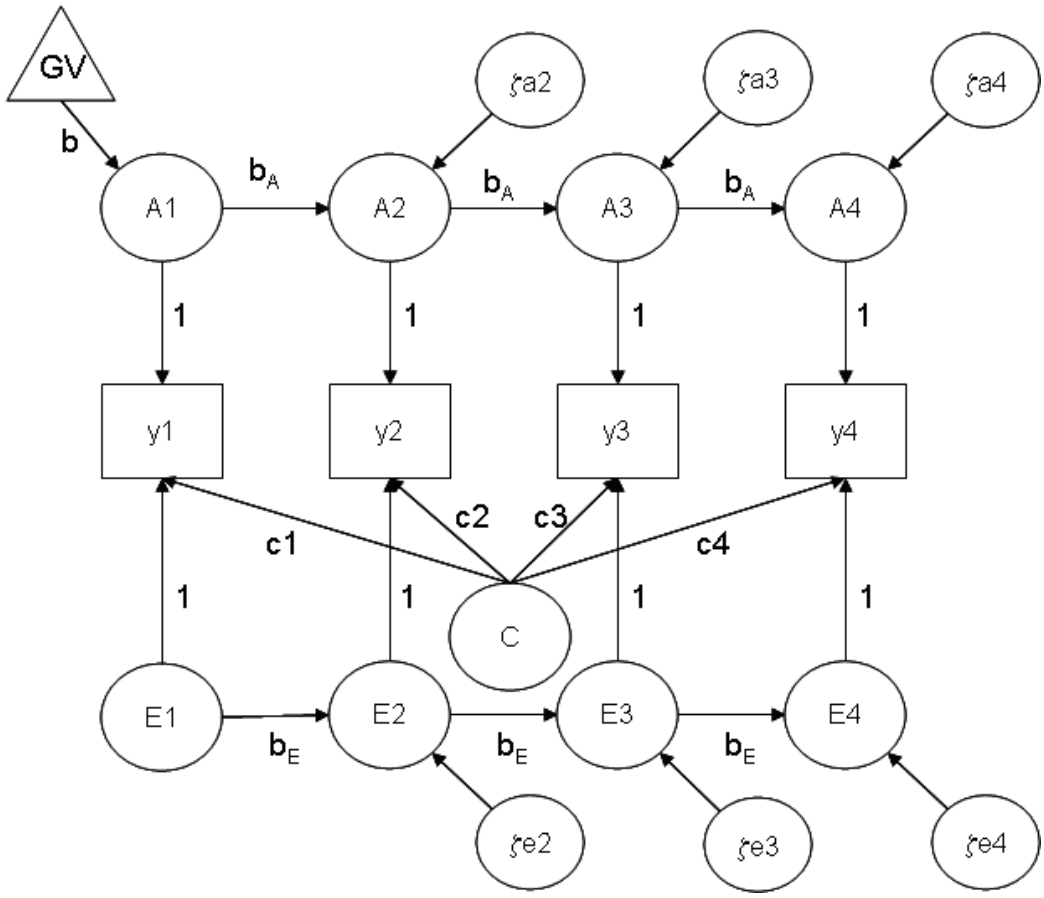

Figure 5

Path diagram for the hybrid simplex-factor model. The triangle represents the GV as fixed regressor. The unit vector, which is included to estimate intercepts is not included to avoid clutter (see Figure 1). In this model the GV enters at occasion 1. In this diagram, the effect size of the GV was defined as .25% of the variance of y1. We also considered the cases in which the GV enters at occasions 2, 3, or 4, and defined the effect size as .25% of the variance of y2, y3, or y4.

factor, and on the autoregressive coefficients in the genetic and unshared environmental simplexes, that is, β_A and β_e. The parameters β_A and β_e equal 1 and .7, respectively. The other parameters in the model are the residual variances, $\sigma^2\zeta_a$ ($\sigma^2\zeta_a$ =.1) and $\sigma^2\zeta_e$ ($\sigma^2\zeta_e$=.204), representing the amount of variance in the genetic factors A(t) that is not explained by the independent factors A(t-1).

The GV was added to the genetic factor (A) at occasion t, and its effect is defined as .25% of the variance in the phenotype y(t) that depends directly on A(t). Because of the genetic autoregression, the GV effect entering at occasion t is transmitted to the phenotypes measured at the subsequent occasions. For example in Figure 5, the GV also affects y_2, y_3 and y_4. The parameter values in the model are given in Table 6, along with the expected phenotypic covariances. As shown in Table 6, the variance due to the common shared environmental factor decreased over time (c^2 decreases from .3 to 0), while the variance due to the genetic factor increased through time (h^2 increases from .3 to .6). We calculated the power in the full multivariate twin model, in the univariate ANOVA of the sum scores, i.e., the sum of the phenotypes observed at the different occasions; in the univariate ANOVA of each individual phenotype, and in the MANOVA. As mentioned above, we did not fit the exploratory factor model on these longitudinal data as the autoregressive covariance structure is not compatible with an exploratory factor model. Table 7 contains the results.

The results are consistent with the results of the preceding studies. First, when the GV affected all phenotypes (enters at occasion 1), the ANOVA of sum scores was the most powerful test of association (p = .71). Its power decreased from .71 to .03, as the GV entered the model at a progressively later occasion. This is expected as the GV signal in the sum score is weakened by the presence of unaffected phenotypes. The power of MANOVA followed a reverse pattern: it was the lowest when all phenotypes were associated with the GV (.51). Given the relatively large phenotypic correlations, this is consistent with Cole at al. (1994; Figure 1; see also Ferreira & Purcell, 2009), and with the results of Study 1 and Study 3. The

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Table 6

Variance Components in Study 4 at the Four Occasions, and the Implied Phenotypic Covariance Matrix, Conditional on GV

Occasion	h^2	c^2	e^2
t_1	.3	.3	.4
t_2	.4	.2	.4
t_3	.5	.1	.4
t_4	.6	0	.4
phenotypic covariance matrix			
1			
.825	1		
.669	.821	1	
.437	.596	.780	1

Note: The model comprises a simplex for A and E, and a common C factor. The factor loadings on the common C factors are √.3, √.2, √.1, and zero. The C factor loading decrease, the additive genetic variances increases, and the unshared environmental variance remained constant. Consequently the total phenotypic variance, conditional on the GV, remains 1 at each occasion. The autoregressive parameters β_A and β_e equal 1 and .7, respectively. The residual variances equal $\sigma^2_{\zeta a}$ =.1 and $\sigma^2_{\zeta e}$=.204

power is high when the GV entered at a later occasion, ranging from .95 (occasion 2) to .88 (occasion 4). The power of the single (affected) phenotype ANOVA had a constant value of .38.

The first column in Table 7 contains the power of the full true multivariate twin model. As in the MANOVA, the power of this model was lowest when the GV entered at occasion 1 (.93). It increased to >.99 when the GV entered at later occasions. The differences in power of the 1 DF test are more pronounced given an alpha of 1E-7: .10 (t1), .85 (t2), .82 (t3), .66 (t4).

Study 5: Stationary Double Simplex (A,E) Model

In the fifth simulation study, we considered a stationary double-simplex model (A, E) with a single phenotype measured at four occasions. Common environmental effects were absent. As in Study 4, the GV was added to the genetic factor (A) at occasion t, and its effect is defined as .25% of the variance in the phenotype y(t) that depends directly on A(t). Due to autoregression, the GV that enters the model at occasion t affects the phenotype at the subsequent occasions. The path diagram of this model is the same as that in Figure 5, except for the absence of C. The genetic autoregression coefficient (β_A) equals .9 and the environmental autoregressive coefficient (β_E) equals .7. The residual genetic and environmental (innovation) variances equal $\sigma^2\zeta_A$ = .114 and $\sigma^2\zeta_E$ = .204, respectively. These parameters resulted in a stationary model, in which the h^2 and e^2 at each occasion equal .6 and .4, respectively. The phenotypic correlations equal .82 (t1-t2, t2-t3, t3-t4), .68 (t1-t3, t2-t4), and .57 (t1-t4). Table 8 contains the results.

The results resemble those of Study 4. We find that the sum score ANOVA is most powerful when all phenotypes were affected (GV entered at the first occasion), and that the power of this ANOVA declines progressively as the GV entered at later occasion. The

Table 7

Power, Non-Centrality Parameters, and Degrees of Freedom (in parentheses) of Univariate and Multivariate Tests of Association Given α = .01 in Study 4

Model	TRUE model	ANOVA Sum scores	ANOVA Single phenotypes at 4 occasions	MANOVA
	2 × 1500	3000	3000	3000
GV at T1	.93,.79	.71	.56 & .56 & .56 & .56	.51
	16.48	9.8	*.38 & .38 & .38 & .38*	10.45
	(1),(4)	(1,2998)	7.51 & 7.51 & 7.51 & 7.51	(4,2995)
			(1,2998)	
GV at T2	>.99,>.99	.41	.01 & .56 & .56 & .56	.95
	40.65	5.51	*.0025 & .38 & .38 & .38*	24.36
	(1),(4)	(1,2998)	0 & 7.51 & 7.51 & 7.51	(4,2995)
			(1,2998)	
GV at T3	>.99,>.99	.15	.01 & .01 & .56 & .56	.93
	39.1	2.45	*.0025 & .0025 & .38 & .38*	23.33
	(1),(4)	(1,2998)	0 & 0 & 7.51 & 7.51	(4,2995)
			(1,2998)	
GV at T4	>.99,>.99	.03	.01 & .01 & .01 & .56	.88
	32.82	.61	*.0025 & .0025 & .0025 & .38*	19.82
	(1),(4)	(1,2998)	0 & 0 & 0 & 7.51	(4,2995)
			(1,2998)	

Note: The power in the true model is included for the likelihood ratio test of the correctly specified GV (1 DF) and for the omnibus test, in which all 4 genetic factors are regressed on the GV (4 DF). In the case of the single phenotype ANOVA, power is reported for α = .01 and α = .01/4 (.0025; 4 phenotypes). The power for the corrected alpha is displayed in italics.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Table 8

Power, Non-Centrality Parameters, and Degrees of Freedom (in parentheses) of Univariate and Multivariate Tests of Association Given $\alpha = .01$ in Study 5

Model	TRUE model	ANOVA sum scores	ANOVA single phenotypes at 4 occasions	MANOVA
	2×1500	3000	3000	3000
GV at T1	.81, .60 12.1 (1),(4)	.52 6.94 (1,2998)	.56 & .45 & .36 & .28 .39 & .29 & .21 & .15 7.51 & 6.08 & 4.93 & 3.99 (1,2998)	.35 7.8 (4,2995)
GV at T2	>.99, >.99 38.91 (1),(4)	.3 4.31 (1,2998)	.01 & .56 & .45 & .36 .0025 & .39 & .29 & .21 0 & 7.51 & 6.08 & 4.93 (1,2998)	.93 23.32 (4,2995)
GV at T3	>.99, >.99 38.73 (1),(4)	.13 2.12 (1,2998)	.01 & .01 & .56 & .45 .0025 & .0025 & .39 & .29 0 & 0 & 7.51 & 6.08 (1,2998)	.93 23.18 (4,2995)
GV at T4	>.99, .99 38.43 (1),(4)	.03 .58 (1,2998)	.01 & .01 & .01 & .56 .0025 & .0025 & .0025 & .39 0 & 0 & 0 & 7.51 (1,2998)	.93 22.94 (4,2995)

Note: The power in the true model is included for the likelihood ratio test of the correctly specified GV (1 DF) and for the omnibus test, in which all 4 genetic factors are regressed on the GV (4 DF). In the case of the single phenotype ANOVA, power is reported for $\alpha = .01$ and $\alpha = .01/4$ (.0025; 4 phenotypes). The power for the corrected alpha is displayed in italics.

power of the MANOVA was lowest when all phenotypes were affected, and increased sharply when the GV entered at a later occasion (see Cole et al. 1994; Ferreira & Purcell, 2009).

The power of the full multivariate twin model resembled that of the MANOVA: the power was relatively low when the GV entered at t1 (.82), but increased sharply when the GV enter at t2 or later (>.99). Given an alpha of 1E-7, the power of the 1 DF test ranges from .03 (GV enters at occasion 1) to .83 (GV enters at occasion 2).

Discussion

In this article, we considered the power of tests of genetic association using multivariate phenotypic data. Our main interest was in power of tests based on sum score ANOVAs, MANOVAs and EFAs in phenotypic data of unrelated subjects. We also reported the power of single phenotype ANOVAs, and the power of likelihood ratio test in the full MZ & DZ twin model.

Based on the results of factor model-based studies (1, 2, and 3), we conclude that overall the EFA is the most powerful model to detect association. The factor model was also found to be powerful to detect linkage by using IBD mapping in sibs (Boomsma, 1996; Boomsma & Dolan, 1998). Medland & Neale (2010) and van der Sluis et al. (in press) also found this approach to be powerful to detect factor level association in single factor models. However, note that in the present paper the success of the EFA in studies 1, 2, and 3 hinges on the fact that the GV effect on the phenotypes is mediated (or conveyed) by common factors, that is, that the factor model is measurement invariant with respect to the GV (van der Sluis et al., in press). This reduces the number of parameters that are estimated to accommodate the mean differences, and so increases the power. Van der Sluis et al. (in press) demonstrated that violation of this invariance (i.e., direct effects of the GV on one or more phenotypes in the model) may greatly reduce the power. We noted that in Studies 1, 2, and 3, the NCP of the MANOVA and the EFA were approximately equal, and the differences in power are solely a function of the number of estimated parameters. Modeling direct effects of the GV on the phenotypes in the EFA (as studied by Medland and Neale, 2010) renders the power of the likelihood ratio test asymptotically equal to the power in the MANOVA.

The power of the MANOVA (and so of the EFA) depends on whether the GV affects all phenotypes, or only a subset, and on the intercorrelations among the phenotypes (as noted by Ferreira and Purcell, 2009). If all phenotypes are affected, the power is relatively low, especially if the phenotypes are relatively highly correlated. If the GV affects a subset of phenotypes, increasing phenotypic correlations can be beneficial. We refer to Cole et al. (1994) for a graphical explanation of these mixed effects.

In the special case of Study 1, the phenotypic sum score is a sufficient statistic, in the psychometric IRT sense: the sum scores contain the same amount of the information as the constituent phenotypic test scores. Under these specific circumstances the tests based on the sum score ANOVA and EFA (subject to measurement invariance) are equally powerful. However, even if the sum score is not sufficient, the sum score ANOVA may still fare well, that is, if the GV effect is present in all phenotypes, as shown in Studies 3, 4,

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

and 5. However, the power decreases with increasing phenotypic correlations, as higher phenotypic correlations result in larger phenotypic (sum score) variance. Also the power in the sum score ANOVA decreases as the variation in the GV effect over the phenotypes increases (Study 3, S31 to S33; see also Medland & Neale, 2010).

The results of the repeated measures studies 4 and 5 are consistent with the results of studies 1 to 3. Specifically, if the GV entered at the first occasion and so affected all phenotypes the power of the MANOVA was relatively low, while the power of the sum score ANOVA was relatively large. The power of the sum scores ANOVA decreased and the power of the MANOVA increased as the GV entered at a later occasion. For instance, as shown in Table 7, when the GV enters occasion 4, the power of the sum score ANOVA and the MANOVA are .03 and .88, respectively. We note that the striking differences in the covariance structures of the repeated measure models (increasing h^2 in Study 4, constant h^2 in Study 5) had little bearing on the power. We did not consider the EFA, as this model is not consistent with repeated measures (Mandys et al., 1994). This is not to say that the factor analytic approach to repeated data is necessarily suboptimal, but the identification of the exact conditions in which an EFA of repeated measures conferred relatively good power to detect a GV is beyond the present scope.

We note the following limitations of the present power study. First, we have chosen configurations of parameter values that we deemed plausible. Many other configurations are possible. For instance, low broad sense heritability (say, .10) does not rule out the presence of quantitative trait loci of relatively large effect. Second, we have limited our analyses to 3 factor models and 2 univariate simplex models. Other models such as multivariate simplex models, or growth curves model may be of interest, depending on the available data. Third, although we reported the power of the true full multivariate twin model, we have made no effort to compare and discuss the power of this model with the power of the other tests ((M)ANOVAs and EFAs), as the study of twins and the study of unrelated subjects differ in sampling requirements (given that about one person in 50 is a twin). In terms of sample sizes, we retain an equal number of cases (3000), but a case in a twin sample naturally consists of two individuals. To arrive at an equal number of individuals, the power in the full twin model could be recalculated for NMZ = 750 and NDZ = 750 using the R code in the Appendix (these results are available on request). However, if twin data are available, genetic association analysis performed in the context of a genetically informative design is very powerful. In addition, the DZ sibpairs provides a within-family test of association that guards against stratification (see Medland and Neale, 2010; Fulker, et al., 1999). Fourth, we have limited our study to multivariate normally distributed data. Multivariate modeling of discrete data is an important issue that remains to be addressed. Fifth, we have limited the phenotypic covariance structure modeling to the exploratory factor model. Confirmatory modeling is often a viable option, is more parsimonious, and may possibly confer greater power. Sixth, in the factor models, the effect of the GV on the phenotypes was conveyed via the common genetic factors. This is in keeping with the notion that a polygenic genetic factor represents the aggregated effects of many loci. However, one cannot discard the possibility that a measured genetic locus may have a direct effect on a given phenotype (see Medland & Neale, 2010, and van der Sluis et al., in press). As studied by Medland and Neale (2010) the GV effect may vary in sign from phenotype to phenotype.

Conclusion

The power studies to date have produced useful information concerning the power to detect the effects of GVs using multivariate data. We note that in the scenarios considered here (see also Medland and Neale, 2010), a multivariate approach is almost always more powerful than a univariate (i.e., single phenotype) approach. However, multivariate data require modeling choices. The reasons for collecting multivariate data depend on the nature of the phenotype(s) of interest (Hottenga & Boomsma, 2008). For instance, if the phenotypes are psychometric indicators, a well fitting common pathway model (e.g., McArdle & Goldsmith, 1984; Neale & Cardon, 1992), or a model involving a single common genetic factor plus relatively small genetic residuals would justify the use of EFA (and in special cases the use of sum scores). However, a set of phenotypes may be viewed as a system of related variables, rather than as a set of psychometric indicators. Huberty and Morris (1989) describe such a system as a 'collection of conceptually interrelated variables that, at least potentially, determine one or more meaningful underlying variates' (p. 304). Clearly this is sufficiently vague to justify the specific advice that one should carry out power analyses tailored to the theoretical and empirical knowledge of the (genetic) covariance structure at hand,[4] rather than rely on general advice.

Acknowledgments

Sophie van der Sluis is financially supported by NWO/MaGW VENI-451-08-025. Dorret I. Boomsma is supported by European Research Council Grant ERC-230374: Genetics of Mental Illness. We thank Manuel Ferreira, Sarah Medland, and Michael Neale for their comments.

Endnotes

1 Ferreira and Purcell (2009) chose the intercorrelations among the phenotypes to be equal, which is consistent with a single factor model.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

2 In exact data simulation, the simulated data fit the true model exactly, and lent themselves to power calculations as the likelihood ratio of the models with and without the GV effect equals the noncentrality parameter of the noncentral χ^2 distribution required to calculate the power.

3 This is part of the MASS library. MVRNORM includes the facility for exact data simulation.

4 The R scripts and Mx scripts used in this study are available on request. These can be tailored to one's own requirements.

References

Allison, D. B., Thiel, B., St. Jean, P., Elston, R. C., Infante, M. C., & Schork, N. J. (1998). Multiple phenotype modeling in gene-mapping studies of quantitative traits: Power advantages. *American Journal of Human Genetics, 63*, 1190–1201.

Amos, C. I., de Andrade, M., & Zhu, D. (2001). Comparison of multivariate tests for genetic linkage. *Human Heredity, 51*, 133–144.

Balding, D. J. (2006). A tutorial on statistical methods for population association studies, *Nature Reviews Genetics, 7*, 781–791.

Birley, A. J., Whitfield, J. B., Neale, M. C., Duffy, D. L., Heath, A. C., Boomsma, D. I., & Martin N. G. (2005). Genetic time-series analysis identifies a major QTL for in vivo alcohol metabolism not predicted by in vitro studies of structural protein polymorphism at the ADH1B or ADH1C loci. *Behavior Genetics, 35*, 509–524.

Boomsma, D. I., & Molenaar, P. C. M. (1987). The genetic analysis of repeated measures I: Simplex models. *Behavior Genetics, 17,* 111–123.

Boomsma, D. I. (1996). Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. *Behavior Genetics, 26*, 161–166.

Boomsma D. I., & Dolan C. V. (2000). Multivariate QTL analysis using structural equation modeling: A look at power under simple conditions. In T. Spector, H. Snieder & A. MacGregor (Eds.), *Advances in Twin and Sib-pair Analysis* (pp. 203–218). London: Greenwich Medical Media Ltd.

Boomsma D. I., & Dolan, C. V. (1998). A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. *Behavior Genetics, 28*, 329–340.

Boomsma, D. I., Geus de, E. J. C., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., Posthuma, D., Beijsterveldt van, T. C. E. M., Hudziak, J. J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics, 9*, 849–857.

Boomsma, D. I., Busjan, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics, 3*, 872–882.

Cole, D. A., Maxwell, S. E., Avrey, R. D., & Salas, E. (1994). How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables. *Psychological Bulletin, 115,* 465–474.

De Geus, E. J. C., & Boomsma, D. I. (2001). A genetic neuroscience approach to human cognition. *European Psychologist, 6,* 241–253.

De Geus, E. J. C., Wright, M. J., Martin, N. G., & Boomsma, D. I. (2001). Genetics of brain function and cognition. *Behavior Genetics, 31*, 489–495.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling: A Multidisciplinary Journal, 16,* 295–314.

Eaves, L. J., Long, J., & Heath, A. C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behavior Genetics, 16,* 143–162.

Evans, D. M., & Duffy, D. L. (2004). A simulation study concerning the effect of varying the residual phenotypic correlation on the power of bivariate quantitative trait loci linkage analysis. *Behavior Genetics, 34*, 135–141.

Evans, D. M. (2008). Factors affecting power and type one errors in association. In B. Neale, M. Ferreira, S. Medland, & D. Posthuma (Eds), *Statistical genetics: Gene mapping through linkage and association* (pp. 487–533). London: Taylor & Francis, Inc.

Ferreira, M. A. R., Visscher, P. M., Martin, N. G., & Duffy, D. L. (2006). A simple method to localise pleiotropic susceptibility loci using univariate linkage analyses of correlated traits. *European Journal of Human Genetics, 14*, 953–962.

Ferreira, M. A. R., & Purcell, S. M. (2009). A multivariate test of association. *Bioinformatics, 25*, 132–133.

Fulker, D. W., Cherny, S. S., Sham, P., & Hewitt, J. K. (1999). Combined linkage and association sib-pair analysis for quantitative traits. *American Journal of Human Genetics, 64*, 259–267.

Gillespie, N. A., Evans, D. E., Wright, M. M., & Martin, N. G. (2004). Genetic simplex modeling of Eysenck's dimensions of personality in a sample of young Australian twin. *Twin Research, 7*, 637–648.

Gordon, D., & Finch, S. J. (2005). Factors affecting statistical power in the detection of genetic association. *The Journal of Clinical Investigation, 115*, 1408–1418.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics, 6,* 95–108.

Hoekstra, R. A., Bartels, M., & Boomsma, D. I. (2007). Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. *Learning and Individual Differences, 17*, 97–114.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Hottenga, J. J., & Boomsma, D. I. (2008). QTL detection in multivariate data from sibling pairs. In B. Neale, M. Ferreira, S. Medland & D. Posthuma (Eds.), *Statistical genetics: Gene mapping through linkage and association* (pp. 239–264). London: Taylor & Francis.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin, 105*, 302–308.

Kettunen, J., Perola, M., Martin, N. G., Cornes, B. K., Wilson, S. G., Montgomery, G. W., Benyamin, B., Harris, J. R., Boomsma, D. I., Willemsen, G., Hottenga, J. J., Slagboom, P. E., Christensen, K., Kyvik, K. O., Sørensen, T. I., Pedersen, N. L., Magnusson, P. K., Andrew, T., Spector, T. D., Widen, E., Silventoinen, K., Kaprio, J., Palotie, A., & Peltonen, L. (2009). GenomEUtwin-project. Multicenter dizygotic twin cohort study confirms two linkage susceptibility loci for body mass index at 3q29 and 7q36 and identifies three further potential novel loci. *International Journal of Obesity, 33*, 1235–1242.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworths.

Mandys, F., Dolan, C. V., & Molenaar, P. C.M. (1994). Two aspects of the simplex model: Goodness of fit to linear growth curve structures and the analysis of mean trends. *Journal of Educational and Behavioral Statistics, 19*, 201–215.

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

Martin, N. G., Boomsma, D. I., & Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics*, *17*, 387–392.

McArdle, J. J., & Goldsmith, H. H. (1984). Structural equation modeling applied to the twin design: Comparative multivariate models of the WAIS. *Behavior Genetics, 14*, 609.

McDonald, R. P. (1999). Test theory: A unified treatment. Mahwah, NJ: Lawrence Erlbaum.

Medland, S. E., & Neale, M. C. (2010). An integrated phenomic approach to multivariate allelic association. *European Journal of Human Genetics, 18*, 233–239.

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

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

Olson, J. M., Witte, J. S., & Elston, R. C. (1999). Tutorial in biostatistics. Genetic mapping of complex traits. *Statistics in Medicine, 18*, 2961–2981.

Peltonen, L. (2003). GenomEUtwin: A strategy to identify genetic influences on health and disease. *Twin Research, 6*, 354–360.

Perola, M., Sammalisto, S., Hiekkalinna, T., Martin, N. G., Visscher, P. M., Montgomery, G. W., Benyamin, B., Harris, J. R., Boomsma, D. I., Willemsen, G., Hottenga, J. J., Christensen, K., Kyvik, K. O., Sørensen, T. I., Pedersen, N. L., Magnusson, P. K., Spector, T. D., Widen, E., Silventoinen, K., Kaprio, J., Palotie, A., & Peltonen, L. (2007). GenomEUtwin Project. Combined genome scans for body stature in 6,602 European twins: Evidence for common Caucasian loci. *PLoS Genetics, 3*, e97.

R Development Core Team. (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Sluis van der, S., Dolan, C. V., Neale, M. C., & Posthuma, D. (2008). Power calculations using exact data simulation: A useful tool for genetic study designs. *Behavior Genetics*, *38*, 202–211.

Sluis van der, S., Verhage, M., Posthuma, D., & Dolan, C. V. (in press). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *Plos-1*.

Spector, T. D., & Williams, F. M.K. (2006). The UK Adult Twin Registry (Twins UK). *Twin Research and Human Genetics, 9*, 899–906.

Vink, J. M., & Boomsma, D. I. (2002). Gene finding strategies. *Biological Psychology, 61*, 53–71.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of heritability for human height. *Nature Genetics, 42*, 565–569.

Willemsen, G., Geus de, E. J., Bartels, M., Beijsterveldt van, C. E., Brooks, A. I., Estourgie van Burk van, G. F., Fugman, D. A., Hoekstra, C., Hottenga, J. J., Kluft, K., Meijer, P., Montgomery, G. W., Rizzu, P., Sondervan, D., Smit, A. B., Spijker, S., Suchiman, H. E., Tischfield, J. A., Lehner, T., Slagboom, P. E., & Boomsma, D. I. (2010). The Netherlands Twin Register biobank: A resource for genetic epidemiological studies, T*win Research and Human Genetics, 13*, 231–245.

Wright, M. J., & Martin, N. G. (2004). Brisbane Adolescent Twin Study: Outline of study methods and research projects. *Australian Journal of Psychology, 56*, 65–78.

https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

Appendix A

R Code for Calculating Power

R code for computing the power of likelihood ratio test statistic. The input are the non-centrality parameter (NCP), the sample size (N), the degrees of freedom (df), and the alpha (alpha). N and alpha can be varied. The actual input in this code is arbitrary.

```
# start power chi2 test
rm(list=ls(all=TRUE)) # wise
powchi=function(alpha,df,NCP) {
cv=qchisq(alpha,df,lower.tail=F)
pchisq(cv,df=df,ncp=NCP,lower.tail=F) }
#
alpha1=.01                  # Input Type I error probability
df=1                        # Input Degrees of freedom
N1=7000                     # Input The sample size N
NCP1=132.6                  # Input NCP
power1=powchi(alpha1,df,NCP1)
print(c(alpha1,NCP1,power1))
N2=3000                     # Input new N
NCP2=N2*(NCP1/N1)
power2=powchi(alpha1,df,NCP2)
print(c(alpha1,NCP2,power2))
alpha2=1E-7                 # Input new alpha
power3= powchi(alpha2,df,NCP2)
print(c(alpha2,NCP2,power3))
# end
```

R code for computing the power of ANOVA. The input are the non-centrality parameter (NCP), the alpha (alpha), the sample size (N). N and alpha can be varied.

```
# start power one way ANOVA (1df test)
rm(list=ls(all=TRUE))       # wise
powanova=function(alpha,df1,df2,NCP) {
cv=qf(alpha,df1,df2,lower=F)
pf(cv,df1,df2,ncp=NCP,lower=F) }
#
alpha1=.01                  # Input type I error probability
N1=3000                     # Input sample size
NCP1=6.94                   # Input non-centrality parameter
df1=1                       # Hypothesis degrees of freedom
df2=N1-2                    # Error degrees of freedom
power1=powanova(alpha1,df1,df2,NCP1)
print(c(alpha1,NCP1,df1,df2,power1))
N2=1000                     # input new N
df2=N2-2
NCP2=(NCP1/N1)*N2
power2=powanova(alpha1,df1,df2,NCP2)
print(c(alpha1,NCP2,df1,df2,power2))
alpha2=.01/4                # input new alpha
power3=powanova(alpha2,df1,df2,NCP2)
print(c(alpha2,NCP2,df1,df2,power3))
```

R code for computing the power of MANOVA. The input are the alpha (alpha), the non-centrality parameter (NCP), the sample size (N), the number of tests (nv), the hypothesis degrees of freedom (df1), the error degrees of freedom (df2). N and alpha can be varied.

```
#start power MANOVA
rm(list=ls(all=TRUE))       # wise
powmanova=function(alpha,df1,df2,NCP) {
fcrit = qf(alpha, df1, df2, lower=F)
```


https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press

```
pf(fcrit, df1, df2, ncp=NCP, lower=F) }
#
alpha1=.01                  # Input alpha
nv=4                        # Input number of tests
N1=3000                     # Input sample size
NCP1=10.45                  # Non-centrality parameter
df1=nv                      # Hypothesis degrees of freedom
df2=N1-nv-1                 # Error degrees of freedom
power1=powmanova(alpha1,df1,df2,NCP1)
print(c(alpha1,NCP1,df1,df2,power1))
N2=6000                     # input new N
NCP2=(NCP1/N1)*N2
df2=N2-nv-1
power2=powmanova(alpha1,df1,df2,NCP2)
print(c(alpha1,NCP2,df1,df2,power2))
alpha2=1E-5                 # input new alpha
power3=powmanova(alpha2,df1,df2,NCP2)
print(c(alpha2,NCP2,df1,df2,power3))
```


https://doi.org/10.1375/twin.13.6.525 Published online by Cambridge University Press