Genetic Epidemiology 11:483–502 (1994)

Inferring the Direction of Causation in Cross-Sectional Twin Data: Theoretical and Empirical Considerations

David L. Duffy and Nicholas G. Martin

Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Australia

A recent multivariate extension of the classical twin study in theory allows the inference of the direction of causation between correlated traits solely using cross-sectional data. In this paper we briefly review this model and assess its usefulness by applying it to a number of pairs of biological and psychological variables between which the nature of the causative relationship is already known. We conclude that the method has a number of biases and limitations. If a causative relationship at the phenotypic level exists between two traits, the correct direction of causation is usually identifiable, providing the reliability and validity of the measures are known. Failure to correctly specify a measurement model can lead to incorrect tests of hypotheses. Difficulties can also occur when discriminating between a direct causative relationship and a correlation due to common genetic or environmental determinants, but these occur in predictable situations. If these considerations are taken into account in interpretation of results, the true nature of the association between traits can often be correctly identified, or at least included in a subgroup of best fitting models. © 1994 Wiley-Liss, Inc.

Key words: twins, causation, smoking, alcohol intake, lung function

INTRODUCTION

A well known shortcoming of cross-sectional and even longitudinal observational studies is that "correlation does not imply causation." To conclude that a given factor causes an outcome usually requires that information from "outside" the study is used to exclude or confirm particular assumptions. These outside sources ultimately rest on experiment. However, human experimentation is limited in the procedures that may be performed by both ethical and logistic considerations. Many potential causes

Received for publication June 15, 1991; revision accepted March 1, 1992.

Address reprint requests to Dr. David L. Duffy, Queensland Institute of Medical Research, 300 Herston Rd., Herston, Queensland 4029, Australia.

of chronic human disease for instance can only be studied in large samples over long periods of time, their effects may not be experimentally increased, and confounding factors cannot be easily controlled.

In recent papers, Heath and coworkers [1989, 1991, 1993] have presented a novel use of the multivariate twin design, to assess the direction of causation between two variables in cross-sectional data using information from the genetic architecture of each trait. Here, the outside information is simply the usual set of assumptions underlying biometrical genetic analysis of the classical twin study.

In the present paper we use this method to examine the association between pairs of physiological and lifestyle variables measured in twins in which the direction of causation between the two variables is known from either longitudinal studies and/or direct experimental manipulations. We have also examined a few cases where the direction of causation can be reasonably inferred from other data. These examples offer a useful practical test of direction of causation methods which have been hitherto largely applied to psychosocial data where the direction of causation is ambiguous or controversial.

METHODS

We shall exemplify the theory behind the methods by examining two simple situations in detail. In path analysis, measured variable A causes measured variable B when a change in the value of A leads (on average) to a proportional change in B, but changes in B do not lead to systematic changes in A [see Bollen, 1989, for a discussion of causation and path analysis]. In the simplest bivariate genetic example using twins (Model 1, see Fig. 1), Trait A is determined by an additive genetic factor and an unshared or random environmental factor (we will denote shared family environment as shared, and restrict the use of the word common to factors influencing more than one trait), while trait B is an imperfect indicator of Trait A (A causes B, which we will write as $A \rightarrow B$), and (since imperfect) affected by its own unshared environment. The relationship between A and B is at the phenotypic level. We will denote models containing this type of relationship as a "directional" model. Then, designating trait A in Twin 1 as $A1$ and so on, h_a^2 the heritability of Trait A, i the coefficient for the path from A to B (and standardising all variables for simplicity), the expected correlation matrices (Σ's) for MZ and DZ twins are:

	MZ twins				DZ twins			
	$A1$	$A2$	$B1$	$B2$	$A1$	$A2$	$B1$	$B2$
$A1$	1				1			
$A2$	h_a^2	1			$\frac{1}{2}h_a^2$	1		
$B1$	i	ih_a^2	1		i	$\frac{1}{2}ih_a^2$	1	
$B2$	ih_a^2	i	$i^2h_a^2$	1	$\frac{1}{2}ih_a^2$	i	$\frac{1}{2}i^2h_a^2$	1.

Under the alternative possibility that B causes A (Model 2, see Fig. 1), i.e., Trait B is an "environmental" determinant of A, then Σ's are:

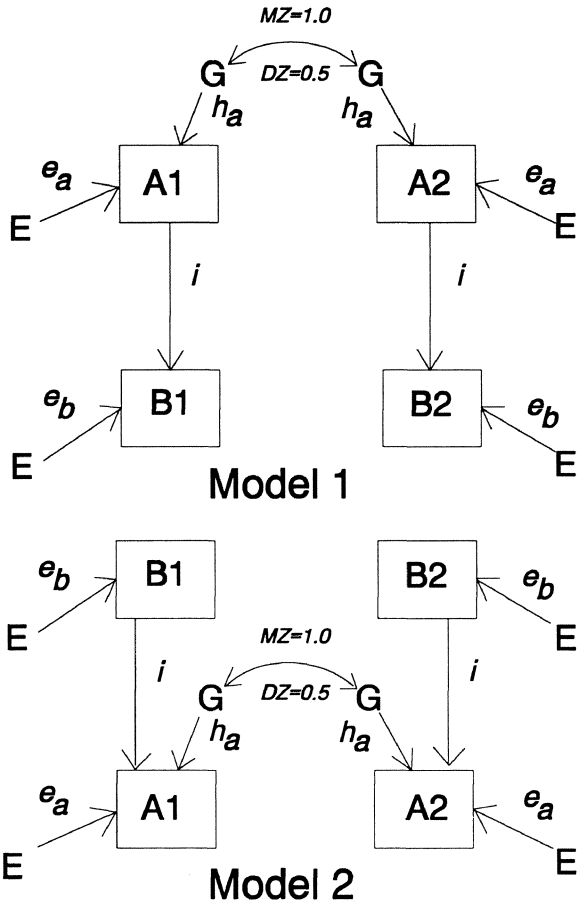

Fig. 1. Path diagrams of the two alternative models of causation where one trait is genetically controlled, the other environmentally controlled, and a phenotypic relationship exists between the two (Models 1 and 2 in the text).

	MZ twins				DZ twins			
	*A*1	*A*2	*B*1	*B*2	*A*1	*A*2	*B*1	*B*2
*A*1	1				1			
*A*2	h_a^2	1			$\frac{1}{2}h_a^2$	1		
*B*1	i	0	1		i	0	1	
*B*2	0	i	0	1	0	i	0	1.

This is equivalent to the model that B is a (perfect or imperfect) indicator of an environmental factor causing both A and B. The main difference between Models 1 and 2 is in their expectation for the cross-twin inter-trait correlations. The heritability of trait B also differs, being $i^2h_a^2$ under the first hypothesis and zero under the second. Unless i and h_a are relatively large, this first term will often be small and do little to resolve the alternative hypothesis.

We now turn to a more realistic case where *A* and *B* each have their own (specific) genetic and unshared environmental causes, but shared environmental, genetic nonadditive, and gene-environment interactive variation are not present (Fig. 2). There are 2^4 possible relationships between *A* and *B*, but only 7 are not overdetermined:

Model 3: *A* and *B* are both caused by independent genetic factors and unshared environmental factors. *A* influences *B* at a phenotypic level.

Model 4: *A* and *B* are both caused by independent genetic factors and unshared environmental factors. *B* influences *A* at a phenotypic level.

Model 5: *A* and *B* are both caused by independent genetic factors and unshared environmental factors. *A* influences *B* at a phenotypic level. *B* reciprocally influences *A* at a phenotypic level.

Model 6: *A* and *B* are genetically correlated. That is an observed correlation between *A* and *B* is due to a common (pleiotropic) genetic factor.

Model 7: *A* and *B* are environmentally correlated. That is an observed correlation between *A* and *B* is due to a common underlying environmental cause (uncorrelated between twins).

Model 8: *A* and *B* are both genetically and environmentally correlated.

Model 9: *A* and *B* are uncorrelated.

It can be shown that Models 5 and 8 make similar predictions for the correlations between twins [Heath, 1993], and in this case are both expected (if the correlation structures for cotwins are equal) to fit perfectly. The other models can be discriminated amongst, given certain conditions. Models 3 and 4, for example, can

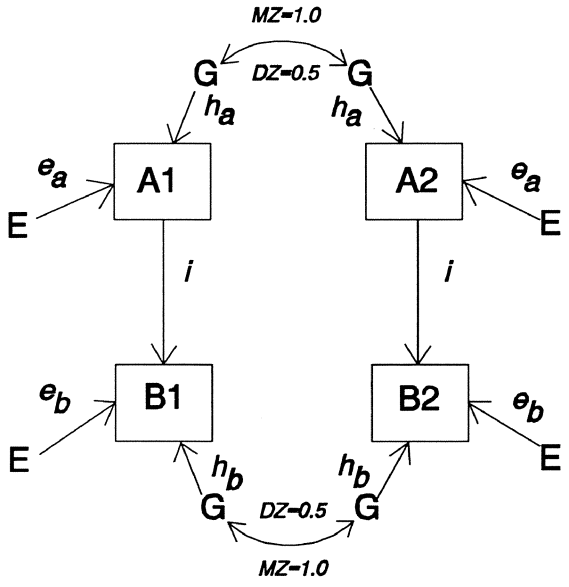

Fig. 2. Path diagram of the model of causation where two traits *A* and *B* are genetically determined, and *A* influences *B* at the phenotypic level (Model 3).

be contrasted as Model 3 predicts Σ's:

	MZ twins				DZ twins			
	$A1$	$A2$	$B1$	$B2$	$A1$	$A2$	$B1$	$B2$
$A1$	1				1			
$A2$	h_a^2	1			$\frac{1}{2}h_a^2$	1		
$B1$	i	ih_a^2	1		i	$\frac{1}{2}ih_a^2$	1	
$B2$	ih_a^2	i	$h_b^2 + i^2h_a^2$	1	$\frac{1}{2}ih_a^2$	i	$\frac{1}{2}h_b^2 + \frac{1}{2}i^2h_a^2$	1

and Model 4:

	MZ twins				DZ twins			
	$A1$	$A2$	$B1$	$B2$	$A1$	$A2$	$B1$	$B2$
$A1$	1				1			
$A2$	$h_a^2 + i^2h_b^2$	1			$\frac{1}{2}h_a^2 + \frac{1}{2}i^2h_b^2$	1		
$B1$	i	ih_b^2	1		i	$\frac{1}{2}ih_b^2$	1	
$B2$	ih_b^2	i	h_b^2	1	$\frac{1}{2}ih_b^2$	i	$\frac{1}{2}h_b^2$	1.

As noted before, the i^2h^2 term will usually be small, so that if h_a and h_b are equal, Model 3 and 4 are equivalent. Both can be differentiated from Model 6, again given favourable conditions. Because a common genetic and specific genetic factors for A and B cannot be estimated simultaneously in the bivariate case, we will parameterize Model 6 as a common additive genetic factor G_c and single specific additive genetic factor G_a. Denoting the coefficients for $G_c = A$ and $G_c = B$ paths as h_{ca} and h_{cb} (Fig. 3), then the Σ's for Model 6 are:

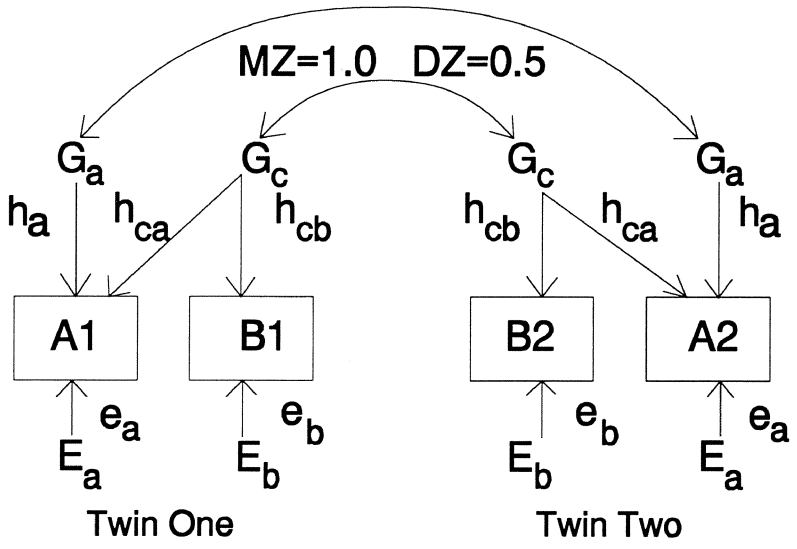

Fig. 3. Path diagram of the model of causation where two traits A and B are genetically correlated (Model 6).

	MZ twins				DZ twins			
	$A1$	$A2$	$B1$	$B2$	$A1$	$A2$	$B1$	$B2$
$A1$	1				1			
$A2$	$h_a^2 + h_{ca}^2$	1			$\frac{1}{2}h_a^2 + \frac{1}{2}h_{ca}^2$	1		
$B1$	$h_{ca}h_{cb}$	$h_{ca}h_{cb}$	1		$h_{ca}h_{cb}$	$\frac{1}{2}h_{ca}h_{cb}$	1	
$B2$	$h_{ca}h_{cb}$	$h_{ca}h_{cb}$	h_{cb}^2	1	$\frac{1}{2}h_{ca}h_{cb}$	$h_{ca}h_{cb}$	$\frac{1}{2}h_{ca}h_{cb}^2$	1.

Therefore, the genetic common factor model constraints the cross-twin intertrait correlations to be equal in the MZ twins, while in the directional models they differ by a ratio of the heritability of the causative trait. It is obvious that as h_a approaches unity and i (or $h_{ca}h_{cb}$) becomes small (so that $ih_a^2 \approx i$, and $i^2h_a^2 \approx 0$) that the $A \rightarrow B$ direction of causation and genetic common factor models become indistinguishable. Practical resolution between these alternatives is difficult with values of i or h_{cb} less than 0.25—the size seen in most of the following practical examples—unless the sample size is large. This is not a problem for the alternative where $B \rightarrow A$, under the same conditions. An interpretation in the case of factor models is that the most heritable trait in the case of genetically correlated traits is the better indicator of the genetic factor causing both traits. Finally, the environmental correlation model (Model 7) might be expected to be the factor model equivalent of a phenotypic causative model. However, the expected cross-twin intertrait correlations are zero.

So far, we have assumed perfect measurement of the variables in the models. A disadvantage of the directional models is that measurement error is not subsumable as unshared environmental variation, as it is in the factor models. If m_a is the correlation between the now unobserved variable A and its observed indicator A', and m_b that for B and B', then the elements of the expected correlation matrix will be attenuated by the following values:

	$A1'$	$A2'$	$B1'$	$B2'$
$A1'$	1			
$A2'$	m_a^2	1		
$B1'$	m_am_b	m_am_b	1	
$B2'$	m_am_b	m_am_b	m_b^2	1.

In this case, if m_a is significantly different from m_b then fitting of directional models to the observed variables A' and B' can lead to incorrect conclusions about the direction of causation. If significant measurement error is present, as will almost inevitably be the case in epidemiological applications, then this must be measured using multiple indicators and/or repeat testing. In the favourable case where both variables have three sources of variation that differ between the traits, then the measurement error variance can be directly estimated for the caused variable [Heath et al., 1993].

Some general comments about the models examined can be made. (1) They differ mainly in their expectations for the cross-twin correlations between traits A and B. (2) The alternative directional models ($A \rightarrow B$ and $B \rightarrow A$) can be differentiated unless they have similar genetic architectures. (3) Measurement error must be explicitly modelled in the directional models or the wrong alternative will be accepted.

(4) The genetic correlation model tends to be confounded with the directional model where the high heritability trait causes the low heritability trait. (5) The environmental correlation model tends to be confounded with the directional model where the low heritability trait causes the high heritability trait. (6) Rejection of the correlation models is easiest when both traits have high heritabilities.

These points are reinforced in Table I, where the results of power calculations for rejection of false models are tabulated. The sample sizes straddle those available to laboratory twin studies and to larger questionnaire based studies. Both these study types are represented in the following practical examples. These come from two data sets previously collected by the second author. Although a number of other multivariate twin studies would have offered larger numbers or stronger associations between variables, most authors do not publish covariance matrices in such a way as to allow this type of analysis. The examples chosen are often more complex than the simple situations discussed above. They include more than two sources of variation, which usually improves the power of the method, as well as covariates such as age.

TABLE I. Probability of Failing to Reject the False Model in a Nested LR χ^2 Comparison With the Reciprocal Causation Model (α = 0.05) When a Directional or a Genetic Factor Model Is the True State of Nature*

	Heritabilities			Number of twin pairs (MZ:DZ = 1:1)		
True Model	h_a^2	h_b^2	Incorrect Model	100	300	900
$A \rightarrow B$	0.1	0.8	Genet factor	0.01	0.00	0.00
(Model 3)			Envir factor	0.94	0.92	0.85
i = 0.25			$B \rightarrow A$	0.05	0.00	0.00
	0.5	0.8	Genet factor	0.09	0.00	0.00
			Envir factor	0.69	0.28	0.01
			$B \rightarrow A$	0.41	0.03	0.00
	0.8	0.8	Genet factor	0.44	0.04	0.00
			Envir factor	0.33	0.01	0.00
			$B \rightarrow A$	0.92	0.87	0.69
	0.8	0.5	Genet factor	0.76	0.43	0.04
			Envir factor	0.34	0.02	0.00
			$B \rightarrow A$	0.76	0.43	0.04
	0.8	0.1	Genet factor	0.85	0.64	0.21
			Envir factor	0.32	0.01	0.00
			$B \rightarrow A$	0.39	0.03	0.00
	0.5	0.5	Genet factor	0.47	0.06	0.00
			Envir factor	0.71	0.32	0.01
Genet factor	0.8	0.1	$A \rightarrow B$	0.70	0.29	0.01
(Model 6)			$B \rightarrow A$	0.16	0.00	0.00
r_g = 0.5	0.8	0.5	$A \rightarrow B$	0.92	0.85	0.66
			$B \rightarrow A$	0.61	0.17	0.00
	0.8	0.8	$A \rightarrow B$ or $B \rightarrow A$[a]	0.28	0.01	0.00
	0.5	0.5	$A \rightarrow B$ or $B \rightarrow A$[a]	0.53	0.10	0.00

*Calculation was done using the non-central χ^2 distribution [see Jöreskog and Sörbom, 1990 for references], and was tabulated by sample size and heritability of traits.

[a]The same probabilities apply to either directional model.

The models were fitted using the structural equation modelling programs LISREL 7.16 [Jöreskog & Sörbom, 1990], and MX [Neale, 1991]. Fitting was performed by maximum likelihood methods in most cases, but used the weighted least squares (WLS) method of Browne [1984; Jöreskog & Sörbom, 1990] in the case of threshold models of ordinal traits. Testing of adequacy of model fit was done by likelihood ratio testing of nested models, and by Akaike Information Criterion (AIC, here $\chi^2 - 2$ (degrees of freedom)).

RESULTS

Data Set 1

Here we examine responses to a mailed questionnaire sent to 5967 adult twin pairs enrolled on the Australian NH&MRC Twin Registry in 1980 [Jardine et al., 1984]. Responses were received from 3808 complete twin pairs. Subjects reported height and weight, alcohol and cigarette use, gynaecological history and a number of health conditions and surgical procedures.

The first example is the relationship between reporting a history of elevated blood pressure (HYP) and being treated with antihypertensive medication (MED). These are strongly correlated (Table II). For the present purposes we will assume that both these items will have high validity and repeatability and fit directly to the observed correlations.

The best fitting bivariate model included an age correction, and additive genetic components for each variable. The most parsimonious model (on the AIC criterion) was the expected HYP → MED (Fig. 4, Table III). However, the MED → HYP model gave a good fit unless the age correlations were constrained to be non-negative, because this model leads to medication use being negatively correlated with age, hypertension positively correlated, and the standardised value of the phenotypic path from medications to HYP was 1.2. The same pattern of age correlations was necessary to allow the genetic factor model to fit as well as it does. Fitting log-linear models to the same contingency tables using GLIM 3.77 found the adjusted correlations between MED and age (4 bands), and HYP and age to be both positive. Zygosity, as one would predict under the HYP → MED model, modified the correlation between twins for HYP, but not MED.

The next (trivial) example shows that the well-known secular positive correlation between year of birth this century and adult height ($r = 0.11$ in females in this group) was not explainable as height causing year of birth (Table IV), as the MZ and DZ correlations for year of birth are by necessity both unity. Therefore an MZ-DZ difference in correlation cannot be induced by height.

TABLE II. Tetrachoric and Biserial Correlations for Age, Hypertension (HYP) and Treatment With Antihypertensive Medication (MED) in 7616 Twins

	Age	HYP	MED
Age	1.00		
HYP	0.42	1.00	
MED	0.68	0.91	1.00

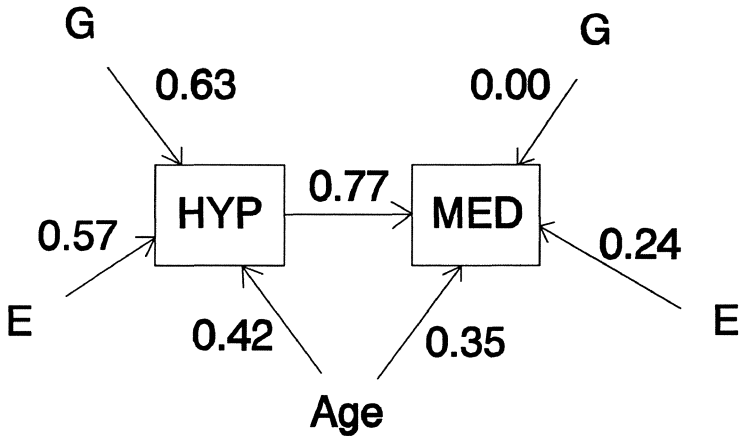

Fig. 4. Preferred direction-of-causation model for hypertension (HYP) and antihypertensive medication use (MED), with age as covariate, in 1799 MZ and 2009 DZ twin pairs—the most plausible model on statistical and other grounds. In the path diagram, all path coefficients are standardised. G represents an additive genetic component, C a shared environmental component and E an unshared environmental component.

TABLE III. Goodness of Fit of Alternative Factor and Direction-of-Causation Models for Hypertension (HYP) and Antihypertensive Medication Use (MED), With Age as Covariate in 1799 MZ and 2009 DZ Twin Pairs

Model	χ^2	d.f.	AIC
Common G, E factor	10.0	14	−18.0
Common G factor	15.5	15	−14.5
Common E factor	44.4	15	+14.4
HYP → MED[a]	**10.0**	**15**	**−20.0**
MED → HYP	12.0	15	−18.0
No correlation	596.0	16	+564

[a]Boldface denotes best fitting models.

TABLE IV. Direction-of-Causation Models for Height and Year of Birth in 1188 MZ and 730 DZ Like-Sex Female Twin Pairs*

Model	χ^2	d.f.	AIC
Reciprocal Model	0.4	7	−13.6
Birth Year → Height[a]	**0.5**	**8**	**−15.5**
Height → Birth Year	30.5	8	+14.5
No correlation	30.6	9	+12.6

*Because year of birth is perfectly correlated between twins, a ridge constant has been added to main diagonal of the observed correlation matrix.
[a]Boldface denotes best fitting models.

TABLE Va. Factor and Direction-of-Causation Models for Final Height and Age at Menarche in 1138 Pairs of MZ Twins and 694 Pairs of DZ Like-Sex Female Twins, Ignoring Reliability of Measurements

Model	χ^2	d.f.	AIC
Common G, E factor	6.6	12	−17.7
Common G factor[a]	**7.0**	**13**	**−19.0**
Common E factor	19.7	13	−6.3
Height → Menarche	**6.9**	**13**	**−19.1**
Menarche → Height	13.4	13	−12.6
No correlation	22.0	14	−6.0

[a]Boldface denotes best fitting models.

A more complex situation is that of age at menarche (in years and months) and final adult height in women. These are known to be positively correlated. By contrast, age at menarche is negatively correlated with height at age 11, and possibly with birth weight and length. Univariate analysis of age of menarche has previously demonstrated a large nonadditive genetic component [Treloar and Martin, 1990]. Both these measures are quite reliable. For a sample of 117 female twins (from data set 2) the correlation between height reported on the questionnaire and that measured in the laboratory was 0.88, while the correlation between recalled age at menarche and actual age has been reported as 0.78 (N = 60) [Damon et al., 1969]. The most reasonable models in this case are that early age of menarche causes decreased height, or that both are coordinated by an underlying genetically controlled trait (a very plausible hypothesis when dealing with body growth/development). Adult height cannot cause the preceding age at menarche directly, and will not be acting as a proxy for height at age 11, in view of the change in sign of correlation with menarche mentioned above. When we first ignored measurement error and fit a specific dominance component for age at menarche and shared environment component to height, the best bivariate models were the genetic factor model and a height to menarche directional model (Table Va, Fig. 5). The heritability of height was higher than that for age at

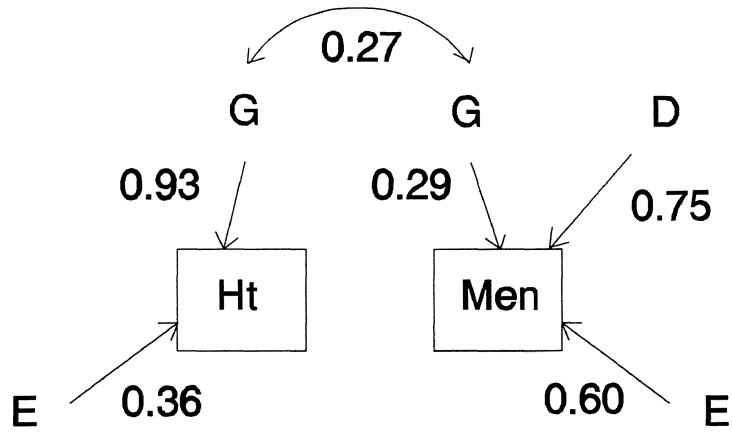

Fig. 5. Preferred direction-of-causation model for final height and age at menarche in 1138 pairs of MZ twins and 694 pairs of DZ like-sex female twins, ignoring reliability of measurements.

TABLE Vb. Direction-of-Causation Models for Final Height and Age at Menarche in 1138 Pairs of MZ Twins and 694 Pairs of DZ Like-Sex Female Twins Incorporating Reliability of Measurement Data*

Model	χ^2	d.f.	AIC
Common G, E factor	15.3	9	−2.7
Common G factor[a]	**15.7**	**10**	**−4.3**
Common E factor	28.5	10	+8.5
Height → Menarche	**15.7**	**10**	**−4.3**
Menarche → Height	20.1	10	+0.1
No correlation	30.9	11	+8.9

*Validation data for height from 117 women, and for age at menarche from 60 women [Damon et al., 1969].
[a]Boldface denotes best fitting models.

menarche, thus explaining the less likely directional model preferred (Type II error). Adding a birth year component to the model (to adjust for secular trend in increasing height and decreasing age at menarche with younger age) does not alter the conclusions. Including the reliability data cited in the model does not alter our conclusions (Table Vb).

Another item on this questionnaire required subjects to report how often the subject had suffered from bronchitis on a four point scale ("never", "only as a child", "occasionally", "frequently"). A history of frequent bronchitis was associated with "Ever" smoking (Tetrachoric r = 0.2; maximum likelihood common (across zygosity and sex) Odds Ratio = 2.1, 95% CI 1.3–2.9, heterogeneity χ^2 not significant). We fitted direction-of-causation models to tetrachoric correlation matrices for the same sex female twins (N_{MZ} = 1204 pairs, N_{DZ} = 714 pairs) using the weighted least squares method. If the reliability of the measures is ignored, the best model on the AIC criterion was the expected smoking → bronchitis directional model though both directional and genetic factor models could not be differentiated on formal LR χ^2 (Table VIa, Fig. 6). It can be argued that the genetic factor model is unlikely on other grounds, of course. While reported "Ever" smoking has a very high reliability, the test-retest tetrachoric correlation for the item on frequent bronchitis in a subgroup

TABLE VIa. Factor and Direction-of-Causation Models for Ever Smoked (SM) and Frequent Bronchitis in 1204 Female MZ and 714 Like-Sex Female DZ Twin Pairs

Model	χ^2	d.f.	AIC
Common G, E factor	6.9	7	−7.1
Common G factor[a]	**7.7**	**8**	**−8.3**
Common E factor	28.3	8	+12.3
Reciprocal pathway	6.9	7	−7.1
SM → Bronchitis	**6.9**	**8**	**−9.1**
Bronchitis → SM	10.3	8	−5.7
No correlation	73.0	9	+55.0

[a]Boldface denotes best fitting models.

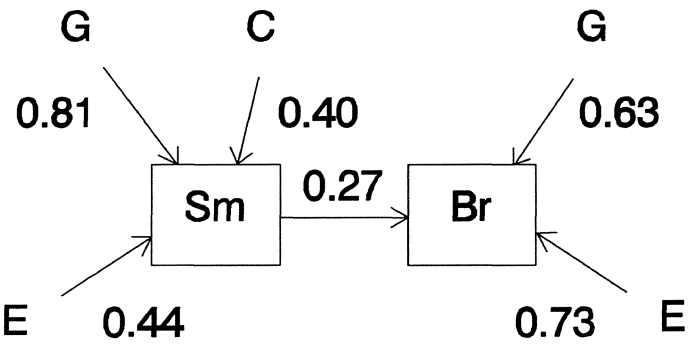

Fig. 6. Preferred direction-of-causation model for Ever Smoked (SM) and frequent bronchitis in 1204 female MZ and 714 like-sex female DZ twin pairs.

(N = 100) retested after 3 months was 0.75 (93% agreement). Incorporating this estimate into the model leads to similar conclusions as above, but the resolving power (because of the small reliability sample) was diminished (Table VIb).

The habit of cigarette smoking has been previously reported to be associated with a number of measurable personality traits, notably extraversion and psychoticism [Eysenck, 1980]. These correlations were present in the data set, most notably for extraversion in women (for females, age-corrected correlation between Ever smoked and Extraversion, 0.20; with Neuroticism, 0.13). One would expect that either extraversion predisposes to cigarette smoking, or both are genetically correlated (in that both exhibit strong genetic components in this data set). Extraversion in females exhibits a nonadditive component in univariate analyses [Eaves et al., 1989]. In the bivariate models fitted for like sex female twins, both expected models were the preferred models (Table VIIa and Fig. 7). If one uses an estimated 8 year test-retest correlation for Extraversion of 0.85 (N = 1400 twins), and a perfect reliability for reported smoking, the same models are again preferred, but again the resolving power was diminished (Table VIIb).

Data Set 2

Here we have reexamined data on alcohol intake, liver enzymes and mean red cell volume (MCV) collected on 80 pairs of same sex male twins with a mean age of 23.5 years [Whitfield and Martin, 1985a, 1985b]. The results for the other 126 pairs

TABLE VIb. Direction-of-Causation Models for Ever Smoked (SM) and Frequent Bronchitis in 1204 Female MZ and 714 Like-Sex Female DZ Twin Pairs*

Model	χ^2	d.f.	AIC
Recip. pathway	5.4	7	−8.6
SM → Bronchitis[a]	**5.4**	**8**	**−10.6**
Bronchitis → SM	7.6	8	−8.4
No correlation	71.5	9	+53.5

*Smoking status taken to be perfectly reliable. Reliability for bronchitis estimated at 0.75 from retest data on 100 individuals.
[a]Boldface denotes best fitting models.

TABLE VIIa. Direction-of-Causation Models for Ever Smoked and Extraversion for 1232 MZ and 751 DZ Like-Sex Female Twins

Model	χ^2	d.f.	AIC
Common G, E factor	3.2	6	−8.8
Common G factor[a]	**3.6**	**7**	**−10.4**
Common E factor	46.8	7	+32.8
Reciprocal pathway	2.3	6	−9.7
Smoking → Extra	6.3	7	−7.7
Extra → Smoking	**2.7**	**7**	**−11.3**
No correlation	97.3	8	+81.3

[a]Boldface denotes best fitting models.

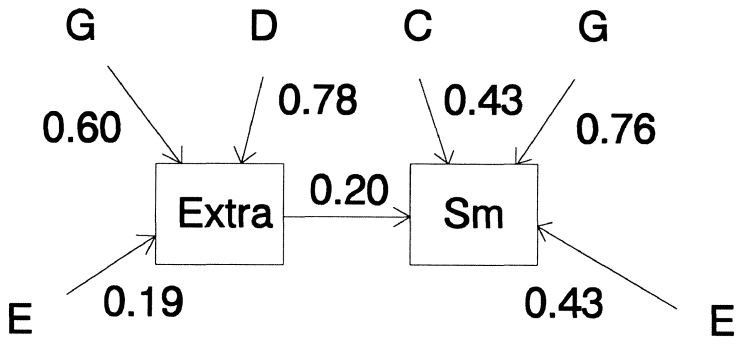

Fig. 7. Preferred direction-of-causation model for Ever Smoked and Extraversion for 1232 MZ and 751 DZ like-sex female twins. In the path model D represents a dominance genetic component.

TABLE VIIb. Direction-of-Causation Models for Ever Smoked and Extraversion for 1232 MZ and 751 DZ Like-Sex Female Twins, Including Age as a Covariate*

Model	χ^2	d.f.	AIC
Comm. G factor	3.6	15	−26.4
Recip. pathway	3.9	14	−24.1
Smoking → Extra	4.9	15	−25.1
Extra → Smoking[a]	**3.4**	**15**	**−26.6**
No correlation	30.3	16	−1.7

*The 8-year reliability of extraversion has been estimated from retesting with a short form scale in 1400 individuals (using the Spearman-Brown formula).

[a]Boldface denotes best fitting models.

TABLE VIIIa. Factor and Direction-of-Causation Models for log Transformed Weekly Alcohol Consumption and Mean Corpuscular Volume (MCV) in 42 MZ and 38 DZ Like-Sex Male Twin Pairs*

Model	χ^2	d.f.	AIC
Common G, E factor	15.8	13	−10.2
Common E factor[a]	**17.2**	**14**	**−10.8**
Common G factor	23.7	14	−4.3
ALC → MCV[a]	16.1	14	−11.9
MCV → ALC	21.6	14	−6.4
No correlation	28.4	15	−1.6

*Measurement error has been ignored.
[a]Boldface indicates best fitting models.

of twins have been discarded as the range of female rates of alcohol use was narrow. Both increased MCV and liver enzyme levels are commonly used in clinical practice as markers of alcohol abuse; longitudinal studies have shown that levels of these markers rise and fall in response to amount of alcohol consumed; and the biological mechanisms mediating the association are well understood. It should be noted that none of the subjects met criteria for alcoholism, and that the range of values seen were all within those seen in healthy populations. The correlations between the plasma enzyme levels—Alanine Aminotransferase (ALT), Gamma-Glutamyl Transpeptidase (GGT), and Aspartate Aminotransferase (AST)—MCV, and reported weekly alcohol intake observed in this study agree well with those previously reported in the literature. [Whitfield and Martin, 1985a]. The liver enzymes and weekly alcohol intake have been (decimal) log transformed to approximate normality. The small sample size means the resolving power of the model tests will be low.

We shall now examine the correlates of alcohol intake in order of strength of association. Looking at the relationship between alcohol intake and MCV (r = 0.30, Table VIIIa and Fig. 8), we found that the true direction of causation gives the best fit out of the directional models, if we ignore measurement error. The common environmental common factor model was preferred over the genetic common factor model.

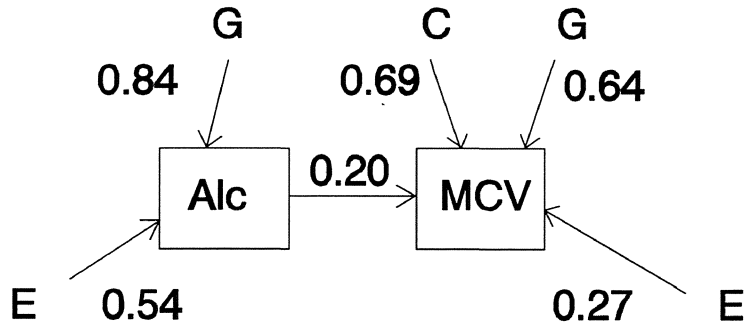

Fig. 8. Preferred direction-of-causation model for log transformed weekly alcohol consumption and mean corpuscular volume (MCV) in 42 MZ and 38 DZ like-sex male twin pairs. Measurement error has been ignored.

TABLE VIIIb. Factor and Direction-of-Causation Models for log Transformed Weekly Alcohol Consumption and MCV in 42 MZ and 38 DZ Like-Sex Male Twin Pairs*

Reliability of measure		Goodness of fit χ^2	
MCV	Alcohol	Alcohol → MCV	MCV → Alcohol
1.0	1.0	7.0	13.4
0.95	0.8	9.3	15.4
0.95	0.7	11.9	15.4
0.95	0.6	14.9	16.5
0.9	0.8	15.8	19.2
0.9	0.7	19.1	20.1
0.9	0.6	20.0	20.1

*Tabulated over a range of reliabilities simulated as coming from an independent sample of 100 individuals tested on two occasions. Models have been fitted to correlation matrices.

We could not discriminate between the directional and the factor model on formal likelihood ratio test, though the AIC was in favour of the directional model.

Estimating the reliabilities for reported alcohol intake and MCV was problematic in this sample. The intraclass test-retest correlation for MCV for a subset of male twins was 0.81 ($N = 48$ individuals). This is just compatible with the lower limit on reliability estimable from the MZ intertwin correlation for MCV of 0.94 (as reliabilities lower than the MZ correlation lead to an estimated true intertwin correlation greater than unity). However, the alcohol item was not repeated on this occasion. A total of 46 pairs did complete two additional alcohol questions on the questionnaire described for Data Set 1. When we fitted a common factor measurement model to these three measures, weekly alcohol consumption correlated 0.82 with the common factor, suggesting a test-retest correlation of 0.67. Use of these estimates, unfortunately, leads to heritabilities (for the true underlying variables) of unity, and model failure (for the Alcohol → MCV model, $\chi^2_8 = 29.5$, $P < 0.01$). As a result, we have chosen to tabulate goodness of fits for the alternative directional models over a range of reliabilities (Table VIIIb).

ALT was the next most highly correlated ($r = 0.21$). Of the two directional models, the correct one gives a superior fit. However, the genetic factor model fits even better.

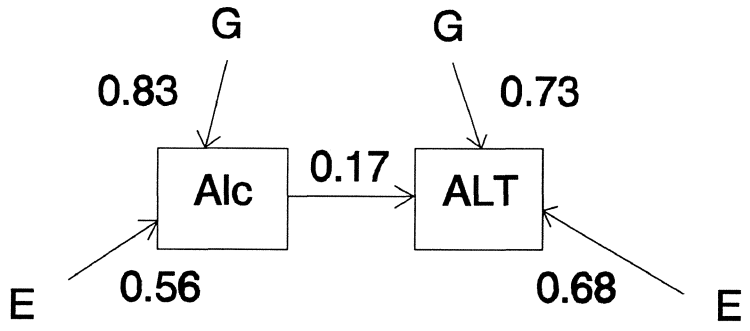

Fig. 9. Preferred direction-of-causation model for log transformed weekly alcohol consumption and log alanine aminotransferase level (ALT) in 42 MZ and 38 DZ like-sex male twin pairs.

TABLE IX. Factor and Direction-of-Causation Models for log Transformed Weekly Alcohol Consumption and log Alanine Aminotransferase Level (ALT) in 42 MZ and 38 DZ Like-Sex Male Twin Pairs

Model	χ^2	d.f.	AIC
Common G, E factor	18.3	14	−9.7
Common G factor[a]	**18.3**	**15**	**−11.7**
Common E factor	26.3	15	−3.7
ALC → ALT	**21.9**	**15**	**−8.1**
ALT → ALC	23.6	15	−6.4
No correlation	26.7	15	−5.3

[a]Boldface denotes best fitting models.

A similar trend was seen for log alcohol consumption and GGT ($r = 0.19$), with the AIC for the ALC → GGT, −13.6, and that for GGT → ALC, −12.4. For AST, the enzyme most weakly correlated with alcohol intake ($r = 0.14$), the correlation between the AST and alcohol intake does not achieve significance ($\chi^2_1 = 3.7$).

Data Set 3

The same subjects also underwent measurement of lung function [Gibson et al., 1983]. The biserial correlation between Forced Expiratory Volume in one second (FEV_1) corrected for height and sex (FEC) and reported current smoking (absent versus present) was 0.16 (higher than that—$r = 0.06$—seen for FEC and number of cigarettes smoked daily). Because the subjects are young and FEV_1 not a sensitive measure of early small airway disease, the correlations are small. The correlations are stronger in females than males. Preliminary univariate analyses found significant heterogeneity of causes of variation of FEC in males and females. In females, a large genetic component was present, but this was absent in males. Therefore we were forced to fit models separately to each sex.

First we will describe the models including current smoking status (SMSTAT) and FEC. Current smoking was a categorical variable, but the sample size did not allow the use of WLS. This leads to inflation of the likelihood ratio χ^2 for models including tetrachoric and biserial correlations, as in this case. Using nested comparisons of model fit could not discriminate between the FEC → SMSTAT and SMSTAT → FEC models in either sex. For pack-years of cigarettes smoked, the association with smoking was only significant in females. No discrimination between the directional models was possible (see Table X, and Fig. 10).

Finally, we present another example of a type 2 error. Height and FEV_1 are strongly correlated (in this data set $r = 0.75$). One would expect this relationship to be a genetic correlation. If an incorrect directional model was picked in this case, we would expect that the path will run from the most heritable to the least heritable trait. For this example, we used all 207 pairs of twins. We found that the genetic factor model would be rejected on conventional criteria (Table XI, Fig. 11), but the height → FEV_1 model fits well. The heritability of height was estimated as 97% and that of FEV_1 as 82%, thus explaining the nature of the "best" model. We can thus predict confounding and allow for this in our conclusions. Both these measures are highly reliable so measurement error will not affect our conclusions.

TABLE X. Factor and Direction-of-Causation Models for Current (Versus Non) Smoking and Age-Height Corrected Forced Expiratory Volume in One Second (FEC) for 44 MZ and 42 DZ Like-Sex Female Twin Pairs Correcting for Age

Model	χ^2	d.f.	AIC
Common G, E factor	18.5	13	−7.5
Common G factor	18.6	14	−9.4
Common E factor	22.9	15	−5.1
Smoking → FEC	19.5	14	−8.5
FEC → Smoking	18.7	14	−9.3
No correlation	25.2	15	−4.8

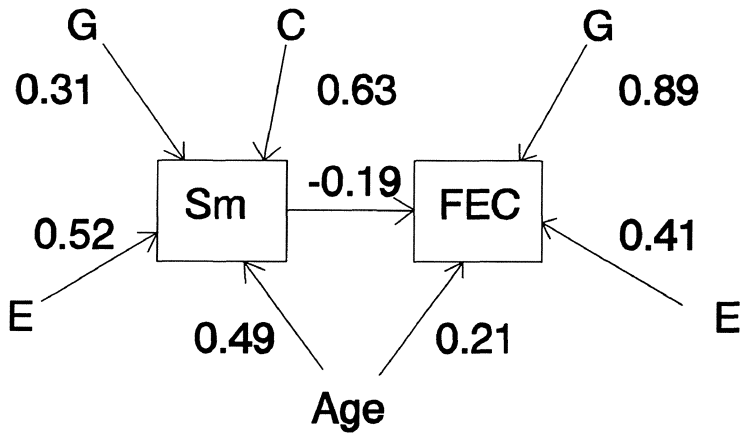

Fig. 10. Preferred direction-of-causation model for current (versus non) smoking and age-height corrected Forced Expiratory Volume in one second (FEC) for 44 MZ and 42 DZ like-sex female twin pairs correcting for age.

TABLE XI. Factor and Direction-of-Causation Models for Height and Forced Expiratory Volume in One Second (FEV-1) in 207 Twin Pairs

Model	χ^2	d.f.	AIC
Common G, E factor[a]	**18.4**	**13**	**−7.6**
Common G factor	38.2	14	+10.2
Common E factor	182	14	+154
Reciprocal pathway	18.4	13	−7.6
Height → FEV_1	**20.2**	**14**	**−7.8**
FEV_1 → Height	56.4	14	+28.4
No correlation	259	15	+229

[a]Boldface denotes best fitting models.

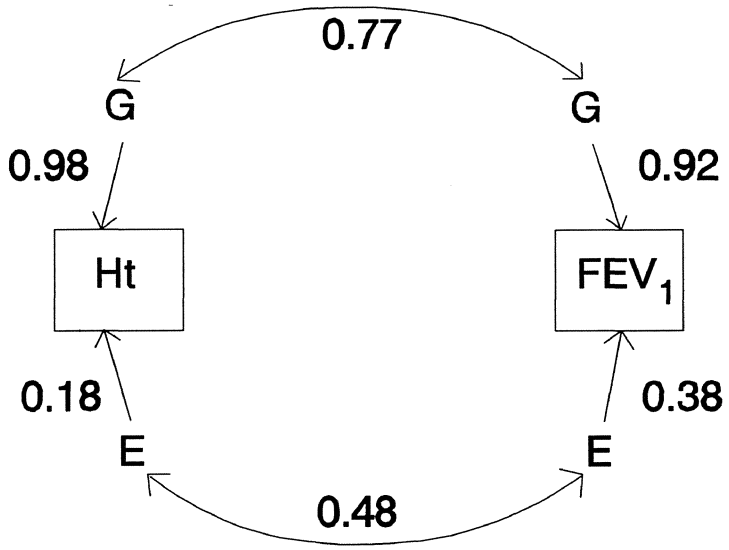

Fig. 11. Preferred factor model for height and Forced Expiratory Volume in one second (FEV_1) in 207 twin pairs.

DISCUSSION

In the first part of this paper, we derived expected covariances for the bivariate case of the classical twin model under a small number of different causative models that included phenotypic causative or genetic and environmental factor models. It was shown that these models gave rise to different expectations, mainly for the cross-twin intertrait correlations, and that in principle temporal precedence and causation could be inferred from cross-sectional data. We then showed that such inferences could be made under ideal circumstances using practical sample sizes. To test the practical use of the method, we then used these models to analyse pairs of correlated traits from previously published data sets where the nature of the association was already known. We chose these examples, as opposed to the use of simulated data sets, to test the robustness of these methods in the face of the various difficulties that can arise in real life. These difficulties include small and inconsistent relationships between variables, the possibility of multiple causal mechanisms underlying associations (including the simultaneous action of nonadditive genetic and shared environmental effect, confounded in the bivariate case). Another difficulty, peculiar to twin studies, is precisely that most of the variables examined seem to have their own specific genetic causes. The method is seen at its best when one variable is genetic, and the other nongenetic.

In the cases where the sample size was small (Data Sets 2 and 3), and the correct model was one of a phenotypic causative relationship, the correct direction out of the two alternatives usually fitted the data better (on the AIC criterion), although the difference between alternatives did not achieve statistical significance on formal likelihood ratio testing. The results in Table I suggest that this would be expected at the given sample sizes and strengths of association. In the large questionnaire data set (Data Set 1), the sample size was sufficient to allow resolution between the al-

ternative directional methods in some cases. Here, the main problem was confounding between the genetic correlation and directional models, and the loss in resolving power due to less than perfectly reliable measures of variables. Where the association between two variables is strong, there is less difficulty differentiating these models. In epidemiology however, effects of major importance in populations are often small when expressed as correlations, as in this case. It is precisely this situation where a non-experimental method for determining direction of causation would be most useful, for example differentiating risk factors from true disease causes. It is unfortunate that in most of the examples we examined, the phenotypic pathway was expected to run from the high to the low heritability trait. As we have noted, this is where these two models are most likely to be confounded.

Therefore, as in path analysis generally, data from other sources must be used to decide whether alternative hypotheses are plausible. For example, it would seem unlikely that a common genetic predisposition to frequent bronchitis and cigarette smoking exists. In the case of extraversion and cigarette smoking by contrast, two hypotheses are plausible. Either the personality type leads to the habit, or both are partly due to (or are indicators for) another genetically controlled trait.

A further point is in the interpretation of causality in these models. It will be quite a common situation for the causal mechanism between two variables to be multistage, and the more remote this relationship, the more likely that a correlational model will be more appropriate than a directional model. In these cases, as noted earlier, one can interpret the "causative" variable in the best fitting directional model as the better indicator of the "true" underlying cause, but this would not fulfil the usual epidemiological criteria for a cause [Rothman, 1986]. Furthermore, in the example of extraversion and cigarette smoking, it is hard to conceptualise mechanistically that a personality trait "causes" the smoking habit per se, in this strict sense, even though we commonly accept the idea that an individual's personality determines his/her actions. This is not the case in the biological examples in data set 2.

There are two main extensions that can be made to these models. The first is to test single or multiple phenotypic causative pathways between observed variables in a multivariate (more than two variable) analysis. The other is to apply them to pairs of latent variables, where multiple indicator traits exist [Heath et al., 1993]. The former, at least, greatly increases the power of tests comparing alternative directions of causation. The question of the nature of the confounding that would occur in the multivariate situation is yet to be examined.

A good deal of enthusiasm has developed about the direction of causation models in the twin and behaviour genetic communities. They have been applied to a number of interesting problems where the question of cause and effect could not be addressed in any other way, though as yet only a few examples have appeared in the literature [Heath et al., 1989, 1991; Duffy and Martin, in press]. We thought it timely to highlight some of the limitations as well as the strengths of these models before they become more widely used. These methods are not infallible or invariably informative, and like model fitting methods, generally require judgement on the part of the user as to their interpretation.

In conclusion, we would highlight that: (1) Failure to allow for measurement error in these models can lead to incorrect hypothesis testing. (2) Unless the association between two traits is strong, a large sample size is required to allow rejection of in-

correct hypotheses using formal likelihood ratio tests. This also applies to validation of measures. (3) When the heritability of trait A is higher than that for trait B then the $A \rightarrow B$ causative model is confounded with the genetic correlation model; while if the environmental determination of A is greater than that of B then the $A \rightarrow B$ causative model is confounded with the environmental correlation model.

ACKNOWLEDGMENTS

The authors thank Michael Neale for providing and giving assistance with MX, and Andrew Heath for much useful advice. D.L.D. is a recipient of an Australian National Health and Medical Research Council (NH&MRC) PHRDC postgraduate medical scholarship. The data sets were collected from members of the Australian NH&MRC Twin Registry, and funded by NH&MRC and Australian Associated Brewers (AAB). The authors also acknowledge the contribution of their collaborators Drs. J. Whitfield, J. Gibson, and J. Oakeshott.

REFERENCES

Bollen KA (1989): Structural equations with latent variables. New York: John Wiley & Sons.

Damon A, Damon ST, Reed RB, Valadian I (1969): Age at menarche of mothers and daughters, with a note on accuracy of recall. Human Biol 41:161–175.

Duffy DL, Macdonald AM, Easton DF, Ponder BAJ, Martin NG (in press): Is the genetics of moliness simply the genetics of sun exposure? A path analysis of nevus counts and risk factors in British twins. Cell Genet Cytogenet.

Eaves LJ, Eysenck HJ, Martin NG (1989): Genes, culture and personality. London: Academic Press.

Eysenck LJ (1980): The causes and effects of smoking. London: Temple Smith.

Gibson JB, Martin NG, Oakeshott JG, Rowell DM (1983): Lung function in an Australian population: contribution of polygenic factors in the Pi locus to individual differences in lung function in a sample of twins. Ann Hum Biol 10:547–556.

Heath AC, Kessler RC, Neale MC, Hewitt JK, Eaves LJ, Kendler KS (1993): Testing hypotheses about direction of causation using cross-sectional family data. Behav Genet 23: 29–50.

Heath AC, Martin NG (1991): Intoxication after an acute dose of alcohol: An assessment of its association with alcohol consumption patterns by using twin data. Alcoholism: Clin Exp Res 15:122–128.

Heath AC, Neale MC, Hewitt JK, Eaves LJ, Fulker DW (1989): Testing structural equation models for twin data using LISREL. Behav Genet 15:9–35.

Jardine R, Martin NG, Henderson AS (1984): Genetic covariation between neuroticism and the symptoms of anxiety and depression. Genet Epidemiol 1:89–107.

Jöreskog KG, Sörbom D (1990): LISREL 7, A guide to the program and applications. Chicago: SPSS Inc.

Neale MC (1991): MX: statistical modelling. Department of Human Genetics, Medical College of Virginia.

Rothman KJ (1986): Modern Epidemiology. Boston: Little, Brown and Co.

Treloar SA, Martin NG (1990): Age at menarche as a fitness trait: Nonadditive genetic variance detected in a large twin study. Am J Hum Genet 47:137–148.

Whitfield JB, Martin NG (1985a): Individual differences in plasma ALT, AST and GGT: Contributions of genetic and environmental factors, including alcohol consumption. Enzyme 33:61–69.

Whitfield JB, Martin NG (1985b): Genetic and environmental influences on the size and number of cells in the blood. Genet Epidemiol 2:133–144.