MR-DoC2: bidirectional causal modeling with instrumental variables and data from relatives

Abstract

Establishing causal associations is an essential step towards developing interventions for psychiatric, substance use and many other disorders. While randomized controlled trials (RCTs) are considered the gold standard for causal inference, they are unethical in many scenarios. Mendelian randomization (MR) can instead be used, but both methods focus on unidirectional causality. In this paper, we developed a new model, MRDoC2, based on the MRDoC model (Minică et al 2018), which includes two polygenic risk scores. It handles bidirectional causation and confounding due to both familial and non-familial sources. OpenMx software was used to study its statistical power, which revealed that there was negligible effect of the difference in sources of variance in the exposure and outcome variables. This finding contrasts with the case for direction of causation modeling with twins but without genotype data, where no power exists when the proportions of trait variance are equal for the two variables.

Keywords: causality, pleiotropy, twin design, mendelian randomization

Luis FS Castro-de-Araujo. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156, P.O. Box 980126, Richmond, VA 23298-0126, USA. Center of Data and Knowledge Integration for Health (CIDACS), Fiocruz, Salvador, Brazil. Dept of Psychiatry, Austin Health. The University of Melbourne, Victoria, Australia. ldearaujo@unimelb.edu.au. orcid:0000-0002-0952-5052

Madhurbain Singh. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156, P.O. Box 980126, Richmond, VA 23298-0126, USA. singhm18@vcu.edu. orcid:0000-0002-9396-2860

Yi (Daniel) Zhou. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156, P.O. Box 980126, Richmond, VA 23298-0126, USA. zhouy33@vcu.edu

Philip Vinh. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156, P.O. Box 980126, Richmond, VA 23298-0126, USA. vinhpb@vcu.edu

Brad Verhulst. Department of Psychiatry and Behavioral Sciences, Texas A&M University. 2900 E 29th Street Bryan Texas, 77802, USA. verhulst@tamu.edu

Conor V Dolan. Department of Biological Psychology, Vrije Universiteit. Amsterdam, Transitorium 2B03, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. orcid:0000-0002-2496-8492

Michael C Neale. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156, P.O. Box 980126, Richmond, VA 23298-0126, USA. Department of Biological Psychology, Free University of Amsterdam. michael.neale@vcuhealth.org. orcid:0000-0003-4887-659X

Introduction

To understand the mechanisms of diseases and disorders, it is essential to differentiate between correlation and causation. Many research designs and current statistical models are insufficient to distinguish correlation from causation. While randomized controlled trials (RCTs) are the gold standard for causal inference (Evans and Davey Smith, 2015), they are unethical or impractical in many common situations. For example, exposure to a traumatic experience, a possible determinant of substance abuse, is a *treatment* that is not amenable to randomization. Mendelian randomization (MR), an instrumental variable method, is an alternative method to evaluate causality when RCTs are not possible. It has improved our understanding of the underlying genetic architecture of several conditions (Ohlsson and Kendler, 2020). However, MR is based on strong assumptions, which can be difficult to validate.

MR is an important method for examining causality in observational studies (Choi et al., 2020; Katikireddi et al., 2018). It uses genetic variants as instrumental variables to detect the causal effect of an exposure on an outcome. As an instrumental variable method, MR depends on instruments that are sufficiently predictive of the exposure. Most mental disorders are complex and highly polygenic, i.e., are subject to the effects of many genetic variants (e.g., SNPs) with small effects on the phenotype of interest. As instruments, genetic variants are likely to be poorly predictive of the exposure, which renders MR susceptible to weak instrument bias. The bias can be in the direction of the confounded observed association, as previously reported (Burgess et al., 2011; Burgess and Thompson, 2013; Hemani et al., 2018). The size and direction of the bias, however, depends on the research study design (see below). The instrument should not share common cause with the outcome and affect the outcome exclusively via the exposure (i.e., no horizontal pleiotropy). Furthermore, the causation is assumed to be unidirectional from exposure to outcome (Hemani et al., 2018; Katikireddi et al., 2018). The *no horizontal pleiotropy* assumption is generally problematic, but particularly in

psychiatry and substance use research, as the relevant phenotypes are characterized by pervasive pleiotropy. For example, variants of CHRNA genes are associated with cigarettes smoked, lung cancer, respiratory phenotypes, and schizophrenia (Minică et al., 2017; Pillai et al., 2009). In two-sample MR, the data originate from two separate genome-wide association studies (GWASs), so no individual has data on both exposure and outcome. By contrast, in one-sample MR genotyped individuals are assessed on both. In one-sample MR, weak instrument bias tends to be in the direction of the confounded observed association, whereas in two-sample MR, the bias is towards the null (i.e., regression dilution; Hemani et al., 2018).

Like RCTs, MR can be applied to infer causality in the presence of confounding, but the approach comes with assumptions that limit its application in some cases. Other designs can be helpful in some cases where MR is not suitable given its assumptions, and we wanted to explore a structural equation model (SEM) that allowed us to complement MR, but allowing full background confounding and reciprocal causation. The Direction of Causation (DoC) model can infer causal relationships by using information from the cross-twin cross-trait correlations, even in the cross-sectional studies (Gillespie et al., 2003; Heath et al 1992; Neale et al., 1994). This requires two phenotypes, which are characterized by appreciable differences between their MZ correlations, or their DZ correlations, or both. However, in order to accommodate bidirectional causality, one must assume that confounding factors are limited to either additive genetic, common environment or specific environmental components of variance. In essence, any three of the five potential sources of covariance between the traits may be estimated, but no more (Figure 1; Duffy and Martin, 1994; Maes et al., 2021).

Minică et al. (2018) combined the classical twin model and the MR model (Heath et al 1993; Gillespie et al., 2003; Neale et al., 1994), in a model for unidirectional causality called MR-DoC. This model accommodates horizontal pleiotropy, i.e., a direct effect of the genetic instrument on the outcome. Following Burgess et al. (2011), Burgess and Thompson (2013), and Kohler et al. (2011), Minică et al. (2018) used a polygenic risk score (PRS) as the instrument, in an attempt to avoid the weak instrument problem associated with single SNPs. Minică et al. (2018) showed that the model has higher power (to detect causality) than traditional MR (Minică et al., 2018). Two strong assumptions of the MR-DoC model are unidirectional causation and the absence of within-person environmental confounding. The latter implies that the correlation between the non-familial environmental influences on the exposure and the outcome is zero. In contrast, confounding originating in genetic and shared environmental correlations can be accommodated in the MR-DoC model.

The present aim is to present an extension to the MR-DoC model, in which we introduce two PRSs, one for each phenotype. Below, we refer to this model as MR-DoC2. Adding the second PRS allows us to model bidirectional causation in the presence of full confounding, originating in shared and unshared environmental and additive genetic effects. The ability to estimate bidirectional causality in the presence of full (shared and unshared environmental and additive genetic) confounding requires the assumption of no horizontal pleiotropy: we assume no direct effect of PRSx on Y or of PRSy on X. However, we note that the model does not imply that PRSx (or PRSy) is uncorrelated with Y (X), we return to this issue below. We first present the model, address the issue of parameter identification, and continue to address the issue of the statistical power.

Methods

All analyses were performed using R version 4.1.1 (R Core Team, 2021). The models were specified using RAM matrix algebra in OpenMx v2.19.6 (Neale et al., 2016). Local model identification implies that the Jacobian matrix of the model is full column rank (Hunter et al., 2021). This was established numerically using the OpenMx function mxCheckIdentification (Hunter et al., 2021). The model is depicted in Figure 2 for an individual twin (rather than a twin pair) to ease presentation. In Figure 2, the two phenotypes (denoted X and Y) are modeled as follows:

$$X_{ij} = a_X A_{Xij} + c_X C_{Xij} + e_X E_{Xij} + g_2 Y_{ij} + b_1 PRS_{Xij},$$

$$Y_{ij} = a_Y A_{Yij} + c_Y C_{Yij} + e_Y E_{Yij} + g_1 X_{ij} + b_3 PRS_{Yij},$$

where *i* denotes twin pair and *j* denotes individual twin. The parameters *ax, cx* and *ex* represent the effects of the additive genetic (*Ax*), shared environmental (*Cx*), and unshared environmental variables (*Ex*) on the phenotype *X* (*ay, cy,* and *ey,* defined analogously as effects of *Ay, Cy,* and *Ey*). The parameters *b1* and *b3* express the instrumental variable effects on *X* and *Y*, respectively. The parameters of main interest are *g1* and *g2,* i.e., the bidirectional causal effects. Note that the model includes correlations between the additive genetic and environmental variables (i.e., cor(*Ax, Ay*)=*ra*, etc.), which are denoted *ra, rc,* and *re.* An important feature of the present model is that it includes the parameter *re.* Finally, note that the instruments may be correlated (correlation *rf* in Figure 2), due to possible linkage disequilibrium.

The vector of parameters is denoted θ = [*ra, rc, re, rf, ax, cx, ex, ay, cy, ey, g1, g2, b1, b3, x, y*], where *x* and *y* estimate the standard deviation of the PRSx and PRSy, respectively. The model's path coefficients were estimated with maximum likelihood estimation, see details of the simulation next section (Tables 2, 3, and S1). Note that the model, as shown in Figure 2, includes the parameters *b2* and *b4*, but these parameters were fixed to zero for identification. However, possible combinations of parameters that render the model identified are presented in Table S1. In particular, one cannot freely estimate *b2* and *b4*, but either can be estimated in the case *re* is equal to zero (Table S1).

Simulation procedure

Statistical power was explored by exact data simulation given various combinations of parameter values (van de Sluis, et al., 2012). We used the R function `mvrnorm` in the R MASS library to simulate exact data (Venables et al., 2002). The power calculation was based on the likelihood ratio (LR) test. Given exact data simulation, parameter estimates of the full (identified) model equal the true values exactly. By subsequently fitting a model with one or more parameters of choice fixed to zero (nested under the true model), we obtain the exact non-centrality parameter (NCP) of the LR test. The NCP is the difference in the LR test statistic under the null and the alternative hypotheses (Verhulst, 2017). Given the NCP, the degrees of freedom (number of parameters fixed to zero), and the chosen Type I error rate alpha (e.g., 0.05), we can calculate the power to reject the hypothesis that the parameters of choice are zero.

The simulation procedure involved: 1) choosing a set of parameter values for the model shown in Figure 2; 2) exact data simulation, with arbitrary N=1,000 MZ pairs and N=1,000 DZ twin pairs; 3) fitting the true model using ML estimation in OpenMx; 4) fitting the false model by fixing one or more parameters to zero and refitting the model; and 5) calculating the NCP and the power to reject the false model restrictions. In the power calculations, we focused on the causal parameters g1, or g2, or g1 and g2 together, i.e., 1 df tests (g1=0 or g2=0), or a 2 df test (g1=g2=0). The 1 df tests are the tests of main interest, because they distinguish unidirectional from bidirectional causation. Note that the 2 df test is also of interest as the full background confounding (ra, rc, and re are all non-zero), implies that X and Y are correlated, regardless of direct causal relations. We investigated the power in a factorial design, in which each parameter featured as a factor, with a given number of levels.

We fitted the true model with all the model parameters freely estimated (and recovered exactly). Next, we fitted three submodels with g1=0, g2=0, and g1=g2=0. The associated NCPs of these three false models were used to calculate the power given the sample sizes of N=1000 MZ and N=1000 DZ and alpha set to 0.05. In addition, we calculated the

R^2 (proportion of explained variance) in the regression of the phenotypes (X and Y in Figure 2) on their respective PRSs, and the correlation between X and Y (Table 4). Given the NCPs and the parameters, we proceeded by regressing the NCPs on the parameter values to gauge the influence of the parameter values on the NCPs (i.e, the statistical power). In this manner, we can determine which parameters, in addition to the parameters g1 and g2, are the most relevant in determining the statistical power. Note that the regression analyses estimate the effect of the parameter values on the NCPs. So they do not directly address statistical power, which depends on the research design, sample sizes, effect sizes, and alpha. To provide some insight into the power given N=1000 MZ and N=1000 DZ, we present power and R^2 for a variety of parameter configurations in Tables 2,3, 4 and S1.

We executed three factorial designs, i.e., combining all chosen parameters' values. The first factorial design included full ACE confounding. Each parameter in the parameter vector θ (except x, y, ex, and ey) was a factor (a total of 13 parameters), each factor had two levels (Table 1), ex was specified as $\sqrt{1-ax^2-cx^2}$ and ey as $\sqrt{1-ay^2-cy^2}$. Therefore, the final number of cells in the first design was 2^{12}=4096 (Table 1 and 2).

The second design was based on an AE model, i.e., we set rc=cx=cy=0. The model without C is of interest, because in adults C influences are generally absent (Polderman et al., 2015). The parameters (i.e., factors) had two or three levels, again ex was specified as $\sqrt{1-ax^2-cx^2}$ and ey as $\sqrt{1-ay^2-cy^2}$. (Table 1 and 3). The number of cells in this design was 3^7*2^2= 8748. Table 4 contains example cells of this design, where the power to reject g1 = 0 (df = 1) was moderate to high.

We executed a third design, again based on an AE model. In this design, we included negative and positive g1 and g2 values, which may be of interest given hypothetical protective (i.e., negative) causal effects (Table S1). In this design, the factors had one or three levels, ex was specified as $\sqrt{1-ax^2-cx^2}$ and ey as $\sqrt{1-ay^2-cy^2}$, and the number of cells in the first design was 4^4=256 (Table 1 and 4). Although we focused on the df = 1 tests throughout, we also report the df = 2 tests in Table S2 for this design.

These factorial designs served to determine the contributions of the various parameters to the statistical power to reject g1=0, g2=0, and g1=g2=0. We selected parameter values such that the phenotypic twin correlations were reasonable (see Table S1) and the predictive strength of the instrument was plausible (Tables 4 and S1). Specifically, we found the highest power with positive b1, b3, g1, and g3 values and the lowest with lowest g1 and g2 values. Tables 2 and 3 show results of linear regression analyses in which we regressed the NCP on the parameter values. The R^2 of each parameter reveals much of the variance in the statistical power is explained by that parameter. Thus, the

contribution of each parameter to the statistical power changes can be quantified and compared.

Results

Power

Based on the simulation results, we identified the following trends concerning the power to reject *g1*=0, *g2*=0 and *g1*=*g2*=0. First, the parameters *b1* and *b3*, i.e. the predictive strength of the genetic instruments, are important to the power. This result can be seen in Tables 2 and 3 by considering the R^2 values of *b1* and *b3*. Specifically, the PRS regression coefficients, *b1* and *b3*, explain around 46% of the variance in the NCP. The causal parameters *g1* and *g2* explain around 40% of the variance in the NCP in the AE models, Table 3. In practice, we found good power (> 0.8, alpha = 0.05, Nmz=Ndz=1000) in a variety of situations with a range of positive and negative values for *b1*, *b3*, *g1*, and *g2* (Tables 4 and S1).

Based on design 2, we found that the correlation between the instruments (*rf*, Figure 2) made little difference to the power, as is clear from the R^2 in Tables 2 and 3. Also, we found that the background correlations, i.e., the *ra, rc* and *re* parameters, are relatively unimportant to the power in the present designs (Table 2 and 3).

We performed two factorial design analyses (with and without C variance), because we wanted to assess how the presence of C variance would affect the power. In Table 2, which includes C variance, *b1* and *b3* explain 36% of the variance of the NCP, which is less than the variance explained in the absence of C variance in the model (namely, 46% R^2 in the AE model, Table 2). So while we see that the parameters related to C have little influence on the power, the absence of C is beneficial. Needless to say, whether C can be dropped in practice will be dictated by the data.

We found that differences in the variance components of *X* and *Y* (i.e., differences in the parameters *ax* vs. *ay* and *cx* vs. *cy*) had little effect on the power. This result is at odds with traditional DoC models, in which the statistical power depends heavily on these differences (Gillespie et al., 2003). Strikingly, in the present model, differences between the phenotypes in variance decomposition is unimportant to the power.

The present results were based on a linear regression of the NCPs on the parameters. We investigated possible non-linearity by including second order polynomials of the parameters. However, we found that the second order terms explained 7.1-8.9% of the variance in the NCPs (g1=0, Tables 2 and 3). From tables 2 and 3 the R^2 ranged from

90.2% (NCP regressed on all other parameters) in the AE model to 94.5% in the ACE model.

In an attempt to put the power analyses in perspective to published results, we present the power curves for four realistic scenarios (Figure 3). The factorial design was performed with values for additive and shared genetic variances based in reports from previous work. The number of twins required to achieve an 80% power of rejecting g1 = 0 is plotted against an increasing R^2 for the PRS of instrument X (path b1). Environmental variances were dependent on shared and additive genetic variances so that *ex* was specified as $\sqrt{1-ax^2-cx^2}$ and *ey* as $\sqrt{1-ay^2-cy^2}$. Additive and shared variances were considered as follows: (A) alcohol use (a^2 49%, c^2 10%) (Verhulst et al., 2015) and heart disease (a^2 22%, c^2 0%) (Wu et al., 2014); (B) BMI (a^2 7.9%, c^2 3%) (Rokholm et al., 2011) and major depression (a^2 37%, c^2 1%) (Scherrer et al., 2003); (C) cannabis use (a^2 51%, c^2 20%) (Verweij et al., 2010) and schizophrenia (a^2 81%, c^2 11%) (Sullivan et al., 2003); (D) dyslipidemia (LDL) (a^2 60%, c^2 28%) (Zhang et al., 2010) and heart disease (a^2 22%, c^2 0%) (Wu et al., 2014) (Figure 3). Figure 3 also includes vertical lines with previously measured estimations of four PRS: Smoking (Pasman et al., 2022), BMI (Furlong and Klimentidis, 2020), LDL (Kuchenbaecker et al., 2019), and attention deficit hyperactivity disorder (ADHD) (Demontis et al., 2019). Figure 3 shows that the power to reject g1=0 is quite reasonable and well within values that appeared in recent publications.

Discussion

We presented a new twin model based on the MR-DoC twin model, which we called the MR-DoC2 model. Compared to the MR-DoC model, MR-DoC2 model accommodates full confounding (originating in background A, C, and E effects) and bidirectional causality. The model can be used to investigate the bidirectional causal interrelationship. As usual in twin studies, the 2x2 phenotypic covariance matrix were decomposed into A, C, and E covariance matrices. However, in addition, phenotypic reciprocal causal paths were specified, as well as paths from the PRSs to phenotypes X and Y.

The MR-DoC2 is suitable for the study of bidirectional causality given A, C, and E confounding, whilst retaining reasonable statistical power. While we have assumed no direct effects of PRSx on Y and PRSy on X, PRSx is correlated with Y and PRSy is correlated with X. There are two sources of these correlations. Given non-zero *rf*, the correlation between PRSx and PRSy, PRSx is correlated with Y via the *rf* path (see Figure 2). The model assumes there are no direct effect of the PRSx on the outcome *Y* (and PRSy on X), but there are indirect effects on Y from PRSx via *rf* (when *rf* not equals zero). As shown in Table S1, there are situations in which the paths b2 and b4 can be

estimated, but these involve other constraints. For instance, the paths b2 or b4 (not both) can be estimated if the unshared environmental correlation *re* is fixed to zero.

As expected, statistical power is higher when a larger proportion of phenotypic variance is explained by the PRS genetic instruments. This is encouraging, because the proportions of trait variance accounted for by SNPs is likely to increase with greater sample sizes and with improvements in the ability to test rare variants' associations with outcomes and their risk factors. We found the sign of the causal parameters (*g1*, *g2*) to be relevant to the power to reject *g1*=0 or *g2*=0. Negative values (*g1*<0, *g2*<0) or a combination of negative and positive values were associated with high power (Table S2). This result is consistent with one previously reported in multivariate twin studies (Evans & Duffy, 2004), where power increases with negative genetic correlation.

In cross-sectional DoC modeling, the statistical power to discern direction of causation is greater when the phenotypes have different variance component proportions. That is to say, if the MZ and DZ correlations for exposure equal those for the outcome variable, the model will not be able to extract information regarding causal direction (Heath et al., 1993). However, in the present article we have shown that the addition of the PRSs permit a larger range of genetic decomposition scenarios and these components have little effect on the ability to detect causal directionality, which is a remarkable advantage over the original DoC model. More importantly, this model is able to fully accommodate A, C, and E confounding, improving on previously noted limitations of DoC models (Rasmussen et al., 2019).

We note that in Tables 2 and 3 the R^2 values did not sum to unity. However, the variances for *b1*, *b3*, *g1*, and *g2* accounts for over 88% of the NCP variance, the remainder are best explained by weak non-linear effects and interactions. We consider these to be too specific and therefore of little interest in conducting a power analysis. All results are based on simulations of 1000 MZs and 1000 DZs, however the code is available in a repository (https://github.com/lf-araujo/mr-doc2), so the reader can perform their own power calculations and include the number of observations that best suits their study.

In the standard DoC model, parameter estimates can be biased when the measurement precision differs between the two variables (Heath et al., 1993). Specifically, the bias in the causal parameters is towards the more reliable variable being the cause of the less reliable one. This effect is modified in the MR-DoC2 model. When reliability is reduced (or error is introduced) the power is reduced as expected, however the estimates of *g1* and *g2* are recovered unbiased, which is an important improvement from DoC.

As seems inevitable when trying to establish causation, some assumptions are necessary. It is important to emphasize that the assumption that horizontal pleiotropy is

absent (i.e., the parameters b2 and b4 are zero) is a strong assumption. Violation of this assumption results in bias in the causal parameters g1 and g2 and in the parameters ra, rc, and re. Specifically, given b2 > 0 and b4 > 0, the parameters g1 and g2 are overestimated as a consequence of fixing b2 = b4 = 0. This overestimation increases the false positive rate, i.e., the rate with which we incorrectly reject g1=0 and/or g2=0. Given b2 > 0 and b4 > 0, the parameters ra, rc, and re are underestimated, as a consequence of the overestimation of g1 and g2.

Although the assumption of no horizontal pleiotropy is undoubtedly a strong one, it is important to note that the present model does include several types of pleiotropy already in the model. First, in Figure 2, the parameter *rf* estimates the correlation between the two PRSs, which may partially or wholly account for association between the PRS for *X* and the outcome *Y*. Second, the bidirectional causation between *X* and *Y* encompasses pleiotropic effects within the causal feedback loop. A useful feature of the MR-DoC2 model is its ability to estimate non-shared (E) confounding, which was considered by some to be a major drawback of standard DoC models (Rasmussen et al., 2019). It seems likely that extending this model to a longitudinal version for repeated measures data would confer additional benefits. Such models are currently being developed in our laboratory for future use.

Multivariate, GxE and DoC approaches were considered less practical because they require larger data sets to obtain accurate estimates of parameters relating two traits (Gillespie et al., 2003). More recently, however, the availability of studies with very many participants have made these models relatively practical to apply. We have shown that the MR-DoC2 model has reasonable power to test bidirectional causation, and therefore is suitable for a range of clinical applications. The next steps in model investigation will be to extend to longitudinal and multivariate data, to provide corroboration of potential causal pathways identified by this modeling.

Table 1. Parameter levels on the three factorial designs, with respective total number of cells for each design simulation. The model specification can be seen in Figure 2. Parameters x and y are not included here, as they remained unchanged in the three designs. Also, *ex* was specified as $\sqrt{1-ax^2-cx^2}$ and *ey* as $\sqrt{1-ay^2-cy^2}$. Table 4 exemplifies rows extracted from Design 3.

θ	Design 1 (ACE)	Design 2 (AE)	Design 3 (AE)
b1	$\sqrt{0.025}$, $\sqrt{0.05}$	$\sqrt{0.025}$, $\sqrt{0.05}$, $\sqrt{0.075}$	$-\sqrt{0.075}$, $-\sqrt{0.03}$, $\sqrt{0.03}$, $\sqrt{0.075}$
b3	$\sqrt{0.025}$, $\sqrt{0.05}$	$\sqrt{0.025}$, $\sqrt{0.05}$, $\sqrt{0.075}$	$-\sqrt{0.075}$, $-\sqrt{0.03}$, $\sqrt{0.03}$, $\sqrt{0.075}$
g1	$\sqrt{0.020}$, $\sqrt{0.05}$	$\sqrt{0.020}$, $\sqrt{0.04}$, $\sqrt{0.06}$	$-\sqrt{0.050}$, $-\sqrt{0.020}$, $\sqrt{0.050}$, $\sqrt{0.020}$
g2	$\sqrt{0.020}$, $\sqrt{0.05}$	$\sqrt{0.020}$, $\sqrt{0.04}$, $\sqrt{0.06}$	$-\sqrt{0.050}$, $-\sqrt{0.020}$, $\sqrt{0.050}$, $\sqrt{0.020}$
ra	*0.25, 0.50*	*0.0, 0.25, 0.50*	0.3
rc	*0.25, 0.50*	*0*	0
re	*0.25, 0.50*	*0.0, 0.25, 0.50*	0.3
rf	*0.25, 0.50*	*0.0, 0.25, 0.50*	0.3
ax	$\sqrt{0.10}$, $\sqrt{0.25}$	$\sqrt{0.10}$, $\sqrt{0.25}$	$\sqrt{0.5}$
ay	$\sqrt{0.10}$, $\sqrt{0.25}$	$\sqrt{0.10}$, $\sqrt{0.25}$	$\sqrt{0.3}$
cx	$\sqrt{0.10}$, $\sqrt{0.25}$	0	0
cy	$\sqrt{0.10}$, $\sqrt{0.25}$	0	0
Total cells	2^{12}=4096	3^7*2^2= 8748	4^4=256

Table 2. Variance explained in statistical power (non-centrality parameter; NCP) by model parameters

	NCP (β^2)		
	g1 = 0	**g2 = 0**	**g1 = g2 = 0**
b1	0.365	0.000	0.181
g1	0.517	0.000	0.289
b3	0.000	0.365	0.181
g2	0.000	0.517	0.289
ra	0.000	0.000	0.000
rc	0.000	0.000	0.000
re	0.002	0.002	0.000
rf	0.041	0.041	0.000
ay	0.002	0.000	0.001
ax	0.000	0.002	0.001
cy	0.002	0.000	0.001
cx	0.000	0.002	0.001
Total R^2	0.929	0.929	0.945

Note: Linear regression of NCP on the parameter values used for 4096 power analyses in the ACE model.

Table 3. Variance explained in statistical power (non-centrality parameter; NCP) by model parameters

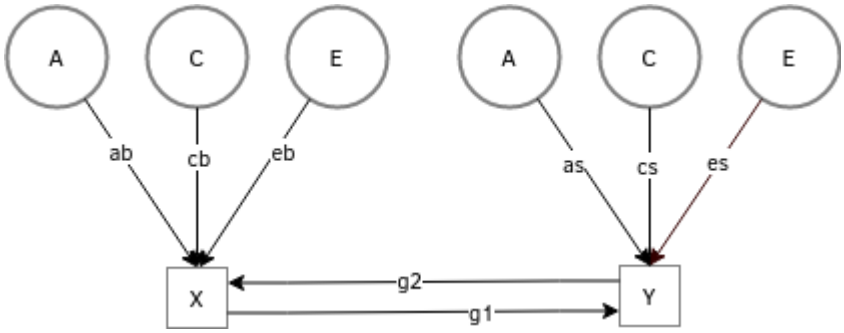
	NCP (β^2)		
	g1 = 0	**g2 = 0**	**g1 = g2 = 0**
b1	0.460	0.000	0.226
g1	0.402	0.000	0.222
b3	0.000	0.460	0.226
g2	0.000	0.402	0.222
ra	0.001	0.001	0.001
rc	0.000	0.000	0.000
re	0.009	0.009	0.001
rf	0.036	0.036	0.001
ay	0.003	0.000	0.002
ax	0.000	0.003	0.002
cy	0.000	0.000	0.000
cx	0.000	0.000	0.000
Total R^2	0.911	0.911	0.902

Note: Linear regression of NCP on the parameter values used for 8748 power analyses in the AE model.

Table 4. Exemplary *b1*, *b3*, *g1* and *g2* values and the power for estimating *g1*. Based from the same factorial AE design presented in Table 3 and parameters set presented in Table 1. Highest power with extreme positive *b1*, *b3*, and *g1* values. Lowest power with lowest g1 and g2 values. Ncpg1, non-centrality parameter for rejecting g1 = 0;powg1, power for rejecting g1 = 0; Y_X, R^2 of the regression of Y on X; Px_X, R^2 of the regression of X on PRSx; Py_Y, R^2 of the regression of Y on PRSy.

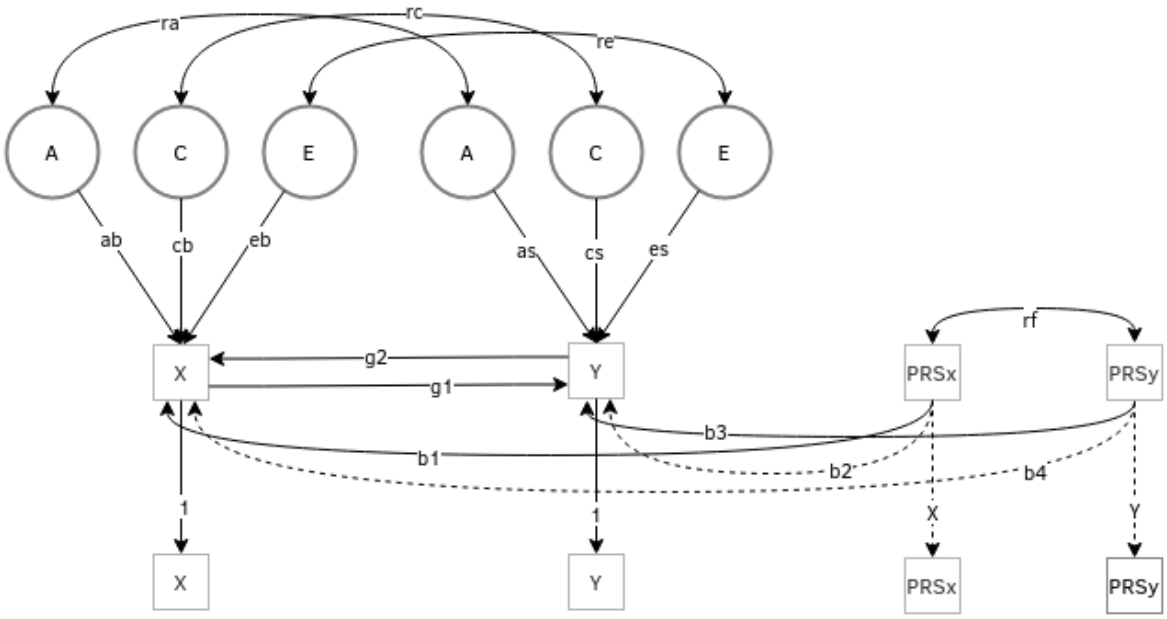
b1	b3	g1	g2	rf	Y_X	Px_X	Py_Y	ncpg1	powg1
0.075	0.075	0.06	0.06	0.00	0.214	0.066	0.066	15.97	0.979
0.075	0.025	0.06	0.02	0.00	0.140	0.068	0.023	15.97	0.979
0.075	0.075	0.06	0.02	0.00	0.138	0.068	0.066	15.97	0.979
0.075	0.050	0.06	0.04	0.00	0.180	0.067	0.045	15.97	0.979
0.075	0.075	0.06	0.04	0.00	0.180	0.067	0.066	15.97	0.979
0.025	0.050	0.06	0.02	0.00	0.334	0.022	0.040	4.71	0.583
0.025	0.075	0.06	0.06	0.00	0.417	0.020	0.059	4.71	0.583
0.075	0.025	0.02	0.02	0.25	0.126	0.070	0.026	4.70	0.582
0.025	0.050	0.02	0.02	0.50	0.465	0.025	0.045	1.13	0.186
0.025	0.050	0.02	0.04	0.50	0.508	0.026	0.045	1.13	0.186
0.025	0.075	0.02	0.02	0.50	0.463	0.026	0.065	1.13	0.186
0.025	0.025	0.02	0.06	0.50	0.539	0.024	0.024	1.13	0.186
0.025	0.075	0.02	0.04	0.50	0.506	0.027	0.065	1.13	0.186

Figure 1. Classic DoC model



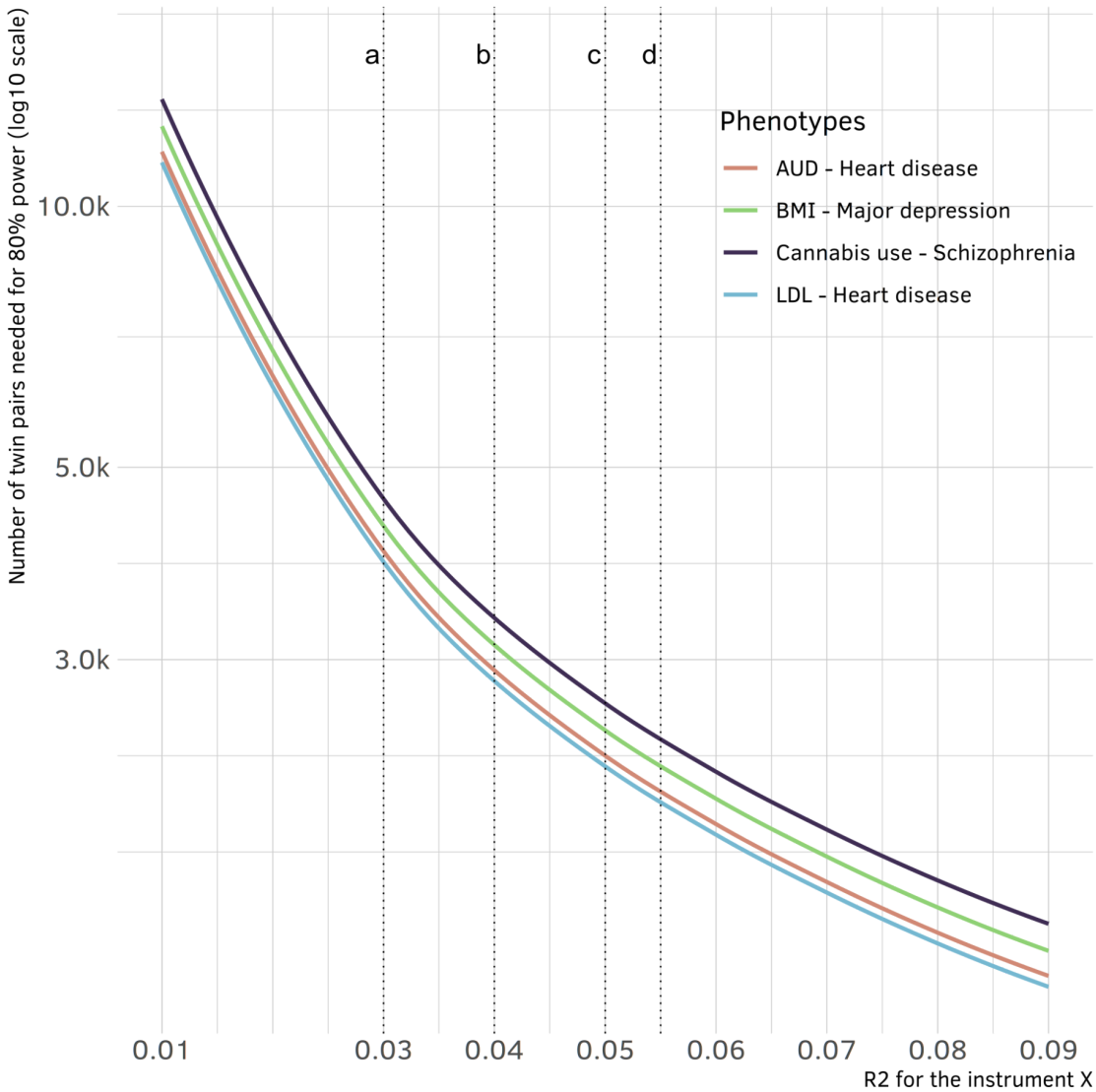
Path diagram representing a Bidirecitonal Direction of Causation model for one twin. This is depicting the relationship between two phenotypes, and the causal paths are estimated affording information from the cross-twin cross-trait correlations. Cross-twin covariance between additive genetic effects is 0.5 (not shown) for DZ twins, as DZs are expected to share 50% of the genetic effects. Standard structural equation modeling symbology is used. Circles represent latent variables. Double-headed arrows are correlations, single-headed are the causal paths. Squares represent observed variables. A, C and E are the additive, shared and unique environmental effects.

Figure 2. MR-DoC2



Path diagram of the MR-DoC2 model for an individual. The model is locally identified when parameters *b2* and *b4* are fixed (zero). The model includes the effects of additive genetic (*A*), common environment (*C*) and specific environment (*E*) factors for both *X* and *Y*, and their effects may correlate (parameters *ra, rc* and *re*). All latent variables have fixed unit variance.

Figure 3. Power curve across R² values for PRSx



Notes: Parameters set for all groups *b3*=*g1*=*g2*=$\sqrt{0.05}$, *ra* = .3; *rc* = .25; *re* = .3; *rf* = .25; Environmental variances were dependent on shared and additive genetic variances: *ex* was specified as $\sqrt{1-ax^2-cx^2}$ and *ey* as $\sqrt{1-ay^2-cy^2}$. Additive and shared variances for the groups: (A) alcohol use (a^2 49%, c^2 10%) (Verhulst et al., 2015) and heart disease (a^2 22%, c^2 0%) (Wu et al., 2014); (B) BMI (a^2 7.9%, c^2 3%) (Rokholm et al., 2011) and major depression (a^2 37%, c^2 1%) (Scherrer et al., 2003); (C) cannabis use

(a2 51%, c2 20%) (Verweij et al., 2010) and schizophrenia (a^2 81%, c^2 11%) (Sullivan et al., 2003); (D) dyslipidemia (LDL) (a^2 60%, c^2 28%) (Zhang et al., 2010) and heart disease (a^2 22%, c^2 0%) (Wu et al., 2014). Vertical lines were added to represent R^2 for four PRSs reported in recent papers: a, smoking (Pasman et al., 2022); b, BMI (Furlong and Klimentidis, 2020); c, LDL (Kuchenbaecker et al., 2019); d, attention deficit hyperactivity disorder (ADHD) (Demontis et al., 2019).

Declarations

Funding

LFSCA is funded by NIH grant no 5T32MH-020030 and the Medical Research Council - UK, Grant no. MR/T03355X/1 during the study.

Conflicts of interest

Authors report no conflicts of interest

Ethics approval

Not applicable

Consent for publication

Not applicable

Availability of data and material

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Code availability

Code is available in a repository for replication.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by **Luis Castro-de-Araujo**, **Madhurbain Singh, Yi (Daniel) Zhou, Philip Vinh, Brad Verhulst, Conor V Dolan** and **Michael C Neale**. The first draft of the manuscript was written by Luis Castro-de-Araujo and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

Burgess, S., Thompson, S.G., 2013. Use of allele scores as instrumental variables for Mendelian randomization. Int. J. Epidemiol. 42, 1134–1144. https://doi.org/10.1093/ije/dyt093

Burgess, S., Thompson, S.G., CRP CHD Genetics Collaboration, 2011. Avoiding bias from weak instruments in Mendelian randomization studies. Int. J. Epidemiol. 40, 755–764. https://doi.org/10.1093/ije/dyr036

Choi, K.W., Stein, M.B., Nishimi, K.M., Ge, T., Coleman, J.R.I., Chen, C.-Y., Ratanatharathorn, A., Zheutlin, A.B., Dunn, E.C., Breen, G., Koenen, K.C., Smoller, J.W., 2020. An Exposure-Wide and Mendelian Randomization Approach to Identifying Modifiable Factors for the Prevention of Depression. Am. J. Psychiatry appi.ajp.2020.19111158. https://doi.org/10.1176/appi.ajp.2020.19111158

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J.I., Grasby, K.L., Grove, J., Gudmundsson, O.O., Hansen, C.S., Hauberg, M.E., Hollegaard, M.V., Howrigan, D.P., Huang, H., Maller, J.B., Martin, A.R., Martin, N.G., Moran, J., Pallesen, J., Palmer, D.S., Pedersen, C.B., Pedersen, M.G., Poterba, T., Poulsen, J.B., Ripke, S., Robinson, E.B., Satterstrom, F.K., Stefansson, H., Stevens, C., Turley, P., Walters, G.B., Won, H., Wright, M.J., Consortium (PGC), A.W.G. of the P.G., Consortium, E.L.& G.E. (EAGLE), Team, 23 and Me Research, Andreassen, O.A., Asherson, P., Burton, C.L., Boomsma, D.I., Cormand, B., Dalsgaard, S., Franke, B., Gelernter, J., Geschwind, D., Hakonarson, H., Haavik, J., Kranzler, H.R., Kuntsi, J., Langley, K., Lesch, K.-P., Middeldorp, C., Reif, A., Rohde, L.A., Roussos, P., Schachar, R., Sklar, P., Sonuga-Barke, E.J.S., Sullivan, P.F., Thapar, A., Tung, J.Y., Waldman, I.D., Medland, S.E., Stefansson, K., Nordentoft, M., Hougaard, D.M., Werge, T., Mors, O., Mortensen, P.B., Daly, M.J., Faraone, S.V., Børglum, A.D., Neale, B.M., 2019. Discovery of the first genome-wide significant risk loci for attention-deficit/hyperactivity disorder. Nat. Genet. 51, 63. https://doi.org/10.1038/s41588-018-0269-7

Duffy, D.L., Martin, N.G., 1994. Inferring the direction of causation in cross-sectional twin data: theoretical and empirical considerations. Genet. Epidemiol. 11, 483–502. https://doi.org/10.1002/gepi.1370110606

Evans, D.M., Davey Smith, G., 2015. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. Annu. Rev. Genomics Hum. Genet. 16, 327–350. https://doi.org/10.1146/annurev-genom-090314-050016

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. Behav. Genet. 34, 135–141. https://doi.org/10.1023/B:BEGE.0000013727.15845.f8

Furlong, M.A., Klimentidis, Y.C., 2020. Associations of Air Pollution With Obesity and Body Fat Percentage, and Modification By Polygenic Risk Score for BMI in the UK Biobank. Environ. Res. 185, 109364. https://doi.org/10.1016/j.envres.2020.109364

Gillespie, N.A., Zhu, G., Neale, M.C., Heath, A.C., Martin, N.G., 2003. Direction of Causation Modeling Between Cross-Sectional Measures of Parenting and Psychological Distress in Female Twins. Behav Genet 33, 14.

Heath, A.C., Kessler, R.C., Neale, M.C., Hewitt, J.K., Eaves, L.J., Kendler, K.S., 1993. Testing hypotheses about direction of causation using cross-sectional family data. Behav. Genet. 23, 29–50. https://doi.org/10.1007/BF01067552

Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., Tan, V.Y., Yarmolinsky, J., Shihab, H.A., Timpson, N.J., Evans, D.M., Relton, C., Martin, R.M., Davey Smith, G., Gaunt, T.R., Haycock, P.C., 2018. The MR-Base platform supports systematic causal inference across the human phenome. eLife 7, e34408. https://doi.org/10.7554/eLife.34408

Hunter, M.D., Garrison, S.M., Burt, S.A., Rodgers, J.L., 2021. The Analytic Identification of Variance Component Models Common to Behavior Genetics. Behav. Genet. 51, 425–437. https://doi.org/10.1007/s10519-021-10055-x

Katikireddi, S.V., Green, M.J., Taylor, A.E., Davey Smith, G., Munafò, M.R., 2018. Assessing causal relationships using genetic proxies for exposures: an introduction to Mendelian randomization. Addict. Abingdon Engl. 113, 764–774. https://doi.org/10.1111/add.14038

Kohler, H.-P., Behrman, J.R., Schnittker, J., 2011. Social Science Methods for Twins Data: Integrating Causality, Endowments, and Heritability. Biodemography Soc. Biol. 57, 88–141. https://doi.org/10.1080/19485565.2011.580619

Kuchenbaecker, K., Telkar, N., Reiker, T., Walters, R.G., Lin, K., Eriksson, A., Gurdasani, D., Gilly, A., Southam, L., Tsafantakis, E., Karaleftheri, M., Seeley, J., Kamali, A., Asiki, G., Millwood, I.Y., Holmes, M., Du, H., Guo, Y., Kumari, M., Dedoussis, G., Li, L., Chen, Z., Sandhu, M.S., Zeggini, E., Group, U.S.S., 2019. The transferability of lipid loci across African, Asian and European cohorts. Nat. Commun. 10. https://doi.org/10.1038/s41467-019-12026-7

Maes, H.H., Neale, M.C., Kirkpatrick, R.M., Kendler, K.S., 2021. Using Multimodel Inference/Model Averaging to Model Causes of Covariation Between Variables in Twins. Behav. Genet. 51, 82–96. https://doi.org/10.1007/s10519-020-10026-8

Minică, C.C., Dolan, C.V., Boomsma, D.I., de Geus, E., Neale, M.C., 2018. Extending Causality Tests with Genetic Instruments: An Integration of Mendelian Randomization with the Classical Twin Design. Behav. Genet. 48, 337–349. https://doi.org/10.1007/s10519-018-9904-4

Minicã, C.C., Mbarek, H., Pool, R., Dolan, C.V., Boomsma, D.I., Vink, J.M., 2017. Pathways to smoking behaviours: biological insights from the Tobacco and Genetics Consortium meta-analysis. Mol. Psychiatry 22, 82–88. https://doi.org/10.1038/mp.2016.20

Neale, M.C., Duffy, D.L., Martin, N.G., 1994. Direction of causation: Reply to commentaries. Genet. Epidemiol. 11, 463–472. https://doi.org/10.1002/gepi.1370110603

Neale, M.C., Hunter, M.D., Pritikin, J.N., Zahery, M., Brick, T.R., Kirkpatrick, R.M., Estabrook, R., Bates, T.C., Maes, H.H., Boker, S.M., 2016. OpenMx 2.0: Extended Structural Equation and Statistical Modeling. Psychometrika 81, 535–549.

https://doi.org/10.1007/s11336-014-9435-8
Ohlsson, H., Kendler, K.S., 2020. Applying Causal Inference Methods in Psychiatric Epidemiology: A Review. JAMA Psychiatry 77, 637–644. https://doi.org/10.1001/jamapsychiatry.2019.3758
Pasman, J.A., Demange, P.A., Guloksuz, S., Willemsen, A.H.M., Abdellaoui, A., ten Have, M., Hottenga, J.-J., Boomsma, D.I., de Geus, E., Bartels, M., de Graaf, R., Verweij, K.J.H., Smit, D.J., Nivard, M., Vink, J.M., 2022. Genetic Risk for Smoking: Disentangling Interplay Between Genes and Socioeconomic Status. Behav. Genet. 52, 92–107. https://doi.org/10.1007/s10519-021-10094-4
Pillai, S.G., Ge, D., Zhu, G., Kong, X., Shianna, K.V., Need, A.C., Feng, S., Hersh, C.P., Bakke, P., Gulsvik, A., Ruppert, A., Lødrup Carlsen, K.C., Roses, A., Anderson, W., Rennard, S.I., Lomas, D.A., Silverman, E.K., Goldstein, D.B., ICGN Investigators, 2009. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet. 5, e1000421. https://doi.org/10.1371/journal.pgen.1000421
Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., Posthuma, D., 2015. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. 47, 702–709. https://doi.org/10.1038/ng.3285
R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
Rasmussen, S.H.R., Ludeke, S., Hjelmborg, J.V.B., 2019. A Major Limitation of the Direction of Causation Model: Non-Shared Environmental Confounding. Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud. 22, 14–26. https://doi.org/10.1017/thg.2018.67
Rokholm, B., Silventoinen, K., Tynelius, P., Gamborg, M., Sørensen, T.I.A., Rasmussen, F., 2011. Increasing Genetic Variance of Body Mass Index during the Swedish Obesity Epidemic. PLoS ONE 6. https://doi.org/10.1371/journal.pone.0027135
Scherrer, J.F., Xian, H., Bucholz, K.K., Eisen, S.A., Lyons, M.J., Goldberg, J., Tsuang, M., True, W.R., 2003. A Twin Study of Depression Symptoms, Hypertension, and Heart Disease in Middle-Aged Men. Psychosom. Med. 65, 548–557. https://doi.org/10.1097/01.PSY.0000077507.29863.CB
Sullivan, P.F., Kendler, K.S., Neale, M.C., 2003. Schizophrenia as a Complex Trait: Evidence From a Meta-analysis of Twin Studies. Arch. Gen. Psychiatry 60, 1187–1192. https://doi.org/10.1001/archpsyc.60.12.1187
Venables, W.N., Ripley, B.D., Venables, W.N., 2002. Modern applied statistics with S, 4th ed. ed, Statistics and computing. Springer, New York.
Verhulst, B., 2017. A power calculator for the classical twin design. Behav. Genet. 47, 255–261. https://doi.org/10.1007/s10519-016-9828-9
Verhulst, B., Neale, M.C., Kendler, K.S., 2015. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. Psychol. Med. 45, 1061–1072. https://doi.org/10.1017/S0033291714002165
Verweij, K.J.H., Zietsch, B.P., Lynskey, M.T., Medland, S.E., Neale, M.C., Martin, N.G., Boomsma, D.I., Vink, J.M., 2010. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies.

Addict. Abingdon Engl. 105, 417. https://doi.org/10.1111/j.1360-0443.2009.02831.x

Wu, S.-H., Neale, M.C., Acton, A.J., Jr, Considine, R.V., Krasnow, R.E., Reed, T., Dai, J., 2014. Genetic and Environmental Influences on the Prospective Correlation Between Systemic Inflammation and Coronary Heart Disease Death in Male Twins. Arterioscler. Thromb. Vasc. Biol. 34, 2168. https://doi.org/10.1161/ATVBAHA.114.303556

Zhang, S., Liu, Xin, Necheles, J., Tsai, H.-J., Wang, G., Wang, B., Xing, H., Li, Z., Liu, Xue, Zang, T., Xu, X., Wang, X., 2010. Genetic and Environmental Influences on Serum Lipid Tracking: A Population-Based, Longitudinal Chinese Twin Study. Pediatr. Res. 68, 316–322. https://doi.org/10.1203/PDR.0b013e3181eeded6

Supplementary material

Table S1. Model Identification. Configurations of fixed and free parameters that may identify the model in Figure 2. Each row shows a combination of fixed or free status for the parameters; the ID column indicates whether the model is locally identified. The text 'fr' indicates that the parameter is freely estimated, and '0' denotes that the parameter is fixed to zero.

ID	x	y	ax	cx	ex	ay	cy	ey	ra	rc	re	rf	b1	b2	b3	b4	g1	g2
FULL																		
no	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr
no	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	fr	fr
yes	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	0	fr	fr
C																		
no	fr	fr	fr	0	fr	fr	0	fr	fr	0	fr	fr	fr	fr	fr	fr	fr	fr
yes	fr	fr	fr	fr	fr	fr	0	fr	fr	0	fr	fr	fr	0	fr	0	fr	fr
E																		
yes	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	0	fr	fr
yes	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	0	fr	fr	fr	fr
no	fr	fr	fr	fr	fr	fr	fr	fr	fr	fr	0	fr	fr	fr	fr	fr	fr	fr
A																		
yes	fr	fr	0	fr	fr	0	fr	fr	0	fr	fr	fr	fr	0	fr	0	fr	fr

Table S2. Exemplary b1, b3, g1 and g2 values and the power for estimating g1. High confounding (ra=re=.3), asymmetric AE (ab =.5, eb =.5, as =.3, es=.7). Highest power with combinations of positive b1, b3, with negative g1 or g2 values. Ncpg1, non-centrality parameter for rejecting g1 = 0;powg1, power for rejecting g1 = 0; Y.X, β^2 of the regression of Y on X. Rmz1 and 2, phenotypic twin correlations.

b1	b3	g1	g2	rf	Y.X	rmz1	rmz2	ncpg1	powg1
0.075	0.030	-0.05	0.05	0.3	0.054	0.578	0.284	12.73	0.946
0.075	0.075	-0.05	0.02	0.3	0.026	0.572	0.340	12.73	0.946
0.075	0.075	-0.05	0.05	0.3	0.072	0.595	0.310	12.73	0.946
0.075	0.030	-0.05	0.02	0.3	0.027	0.574	0.294	12.73	0.946
0.075	0.030	-0.05	-0.05	0.3	0.033	0.605	0.347	12.73	0.946
0.075	0.075	-0.02	0.02	0.3	0.080	0.598	0.334	5.00	0.609
0.075	0.030	-0.02	0.02	0.3	0.076	0.594	0.293	5.00	0.609
0.075	0.030	0.02	-0.05	0.3	0.043	0.501	0.356	4.23	0.538
0.030	0.030	0.02	0.05	0.3	0.328	0.634	0.396	1.69	0.255
0.030	0.030	0.02	0.05	0.3	0.328	0.634	0.396	1.69	0.255
0.030	0.030	0.02	0.05	0.3	0.343	0.642	0.401	1.69	0.255
0.030	0.030	0.02	-0.05	0.3	0.034	0.468	0.349	1.69	0.255
0.030	0.075	0.02	0.05	0.3	0.343	0.647	0.451	1.69	0.255