

## **Twins and Causal Inference: Leveraging Nature's Experiment**

### **Authors:**

T. A. McAdams<sup>1,2</sup>, F. V. Rijsdijk<sup>1</sup>, H. M. S. Zavos<sup>3</sup>, J-B Pingault<sup>1,4</sup>

### **Affiliations**

<sup>1</sup> Social, Genetic & Developmental psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London

<sup>2</sup> Promenta Research Centre, University of Oslo

<sup>3</sup> Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London

<sup>4</sup> Clinical, Educational and Health Psychology, Division of Psychology and Language Sciences, University College London.

Correspondence: Tom A. McAdams, Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, De Crespigny Park, London, SE5 8AF. Email: [tom.mcadams@kcl.ac.uk](mailto:tom.mcadams@kcl.ac.uk)

### **Acknowledgements**

T.A.M. is supported by a Sir Henry Dale Fellowship, jointly funded by the Wellcome Trust and the Royal Society (107706/Z/15/Z). J.B.P. is supported by an MQ Transforming Mental Health (MQ16IP16) award and the Medical Research Foundation 2018 Emerging Leaders 1<sup>st</sup> Prize in Adolescent Mental Health.

## Abstract

In this paper we discuss how samples comprising monozygotic and dizygotic twin pairs can be used for the purpose of strengthening causal inference by controlling for shared influences on exposure and outcome. We begin by briefly introducing how twin data can be used to inform the biometric decomposition of population variance into genetic, shared-environmental and non-shared environmental influences. We then discuss how extensions to this model can be used to explore whether associations between exposure and outcome survive correction for shared aetiology (common causes). We review several analytical approaches that can be applied to twin data for this purpose. These include multivariate structural equation models, co-twin control methods, direction of causation models (cross-sectional and longitudinal), and extended family designs used to assess intergenerational associations. We conclude by highlighting some of the limitations and considerations that researchers should be aware of when using twin data for the purposes of interrogating causal hypotheses.

## Introduction

Twin studies are most often associated with the estimation of heritability and understanding the extent to which correlated traits share common genetic causes. This focus on genetics often means that the role of twin studies in understanding non-genetic influences on human behaviour is overlooked. However, twin studies have demonstrated that around half of the variance in human traits is attributable to non-genetic influences (Polderman et al., 2015). Furthermore, as well as partitioning variance into genetic and environmental influences, it is possible to use twin studies to ask whether associations between exposure and outcome remain after accounting for the confounding effects of shared aetiology (common causes). In this manner, twin studies can be thought of as quasi-experimental. By controlling for common causes twin studies can test what have been referred to as “quasi-causal hypotheses” (Turkheimer & Harden, 2014). The use of phrases like *quasi-causality* and *causal inference* (as opposed to straightforward *causality*) highlight that twin studies (like other epidemiological approaches) cannot typically demonstrate causality between two variables – they cannot prove that X (the exposure) *causes* Y (the outcome). They can however show that X does (or does not) continue to predict Y after controlling for shared aetiology, thereby strengthening the ability of researchers to make causal inferences. This is the primary focus of our review.

## The Twin Model

Before discussing the use of twin studies in causal inference it is necessary first to briefly introduce the logic underlying twin studies. The classical twin design typically decomposes phenotypic variance ( $V_p$ ) into three components: additive genetic (A), shared environment (C) and non-shared environment (E) influences. Additive genetic influences (A) refer to effects of alleles or loci which act in an additive manner (two copies of a risk allele at the same locus confer twice the risk of one copy)<sup>1</sup>. Shared environment

---

<sup>1</sup> It is possible to model dominant genetic effects in twin models (typically denoted by a D), where effects of risk alleles are modelled as multiplicative rather than additive. However, a typical twin dataset only provides enough information (degrees of freedom) to decompose variance into 3 components, so researchers need to choose between an ACE or an ADE model. Covariance patterns inform this choice: If MZ correlations are more than twice the size of DZ correlations, then an ADE model is appropriate. If DZ correlations are more than half the size of MZ correlations, then an ACE model is appropriate. ACE models tend to be fitted more often than ADE models, but all concepts discussed in this paper apply to both. Some extended family designs/datasets allow for ACDE models to be fitted.

(C) refers to environmental effects which make members of the same family similar to one another. In twin studies the shared environment is typically defined as environmental effects that make twins similar to one another regardless of zygosity.<sup>2</sup> Because twins share fewer environmental experiences as they age, the effect of the shared environment diminishes with age for most traits (e.g. Haworth et al., 2010). In extended family studies it is possible to vary the definition of shared environmental effects to include e.g. siblings or entire nuclear families. The non-shared environment (E) refers to effects which make members of the same family different from one another. While typically referred to as the non-shared *environment*, the E component in a twin model comprises the residual variance in a trait remaining after genetic and shared environmental effects are considered. As such, E will comprise error variance and the effects of random non-shared biological events (Smith, 2011; Tikhodeyev & Shcherbakova, 2019), as well as true non-shared environmental influences.

In Figure 1 we present the classical univariate twin model used to decompose variance in a trait into A, C and E components. The method involves comparing the degree of resemblance between pairs of monozygotic (MZ) twins and dizygotic (DZ) twins. Resemblance within MZ twin pairs ( $r_{MZ}$ ) is due to twins sharing all genes and a shared environment ( $r_{MZ} = A + C$ ). Resemblance within DZ pairs ( $r_{DZ}$ ) is expressed as  $r_{DZ} = \frac{1}{2}(A) + C$ ; as DZ twins share only half their segregating genes on average but they are assumed to share the shared environment to the same extent as MZ pairs (100%).

In the twin design, heritability is calculated as twice the difference between MZ and DZ correlations:  $A = 2(r_{MZ} - r_{DZ})$ . Shared environment can be estimated as the difference between the MZ correlation and the heritability:  $C = r_{MZ} - A$ . As non-shared environment is the only influence which acts to make MZ twins different from one another, it can be calculated as the total phenotypic variance (often standardised to 1 for ease of interpretation) minus the MZ correlation:  $E = V_p - r_{MZ}$ .

INSERT FIGURE 1 AROUND HERE

### Understanding Confounding using Twin Models

In their “Top 10 replicated findings from behavioural genetics”, 2 of the most replicated findings highlighted by Plomin, DeFries, Knopik, and Neiderhiser (2016) are of crucial importance to researchers interested in drawing causal inference from associations between variables in non-experimental studies:

1. *Phenotypic correlations between traits often show substantial genetic influence.*

In other words, when 2 traits correlate, it is often because they share common genetic causes. This can be attributed to the joint influences of pleiotropy (whereby genetic variants associate with multiple traits) and polygenicity (whereby traits are associated with many genetic variants). The possibility of shared genetic influences highlights a major reason why associations between traits should not be interpreted as indicative of causal influences of one trait on another. Importantly, this applies to prospective as well as cross-sectional associations. That is, we cannot assume that genetic factors that influence trait X (exposure) at time point 1 will not influence trait Y (outcome) at time

---

<sup>2</sup> This constitutes the Equal Environments Assumption (EEA): the assumption that environmental influences that operate on a trait are shared equally within MZ and DZ twin pairs. If environmental influences that make twins alike are more similar for MZ pairs than DZ pairs—for reasons besides MZ pairs being more genetically alike—then twin model estimates will be biased. Studies of misclassified twins (MZ pairs labelled as DZs and vice versa) provide a powerful test of the EEA and lend support to the validity of this assumption (Conley, Rauscher, Dawes, Magnusson, & Siegal, 2013; Kendler, Neale, Kessler, Heath, & Eaves, 1993).

point 2, and thus explain some or all of their covariance. For example, anxiety in childhood is predictive of later depression, and while this may be because anxiety leads to depression, it is also possible that genetic influences underlying childhood anxiety may manifest as depression later in life (Rice, van den Bree, & Thapar, 2004).

It is worth considering here that the above discussion assumes *horizontal* pleiotropy, whereby genetic variants influence both X and Y and induce or inflate covariance between them. However, genetic covariance between X and Y can also come about through *vertical* pleiotropy, whereby X causally influences Y, resulting in genetic variants associated with X becoming associated with Y. Horizontal pleiotropy constitutes a source of confounding on the association between X and Y, whereas vertical pleiotropy does not. In practice it is difficult to distinguish these two possibilities, and most twin studies aimed at testing quasi-causal hypotheses do not attempt to do so, instead accounting for genetic covariance between traits and evaluating whether associations persist. Where genetic covariance is attributable to vertical pleiotropy, associations may therefore be *overcorrected*. However, provided that genetic factors do not explain *all* of the variance in X, then controlling for vertical pleiotropy as though it were horizontal pleiotropy will not reduce the association between X and Y to zero, because the residual variance in X will still predict Y. Given that no phenotypes are 100% heritable (or indeed 100% familial; Plomin et al., 2016; Polderman et al., 2015), then where a significant causal effect of X on Y exists, their association should not reduce to zero when controlling for potential familial confounds using twin methods. Implicit here is the expectation that the residual covariance will be large enough to be detected. A caveat to this is that where heritability of X is very high, then power to detect a residual association after accounting for genetic correlation may be limited.

## 2. *Most measures of putative 'environmental exposures' show genetic influence.*

Measures that index experiences and environments such as life events, social support, and exposure to toxins (i.e. variables commonly conceptualised as environmental exposures that may influence human traits) correlate with (and may be under the influence of) genetic factors (Jaffee & Price, 2007; Kendler & Baker, 2007). This is referred to as gene-environment correlation (rGE) and has important consequences for researchers interested in drawing causal inferences from associations between putative environmental exposures and outcomes. That is, if an environmental exposure is predictive of a particular outcome, but both outcome and exposure are heritable, then it is possible that the association is partially or entirely accounted for by shared genetic influences. Again, the joint influences of pleiotropy and polygenicity mean that genetic correlations between putative environmental exposures and outcomes should be considered likely.

Three forms of rGE are typically distinguished from one another and give some insight into why putative environmental variables may be heritable: **1. Passive rGE** occurs when people are born into an environment that correlates with their genotype. For example, most children are reared by their biological relatives, so many elements of their environment will be correlated with their genotype. Recently, this concept has been reintroduced into the literature in discussions surrounding “genetic nurture” and “dynastic effects” – the environmentally mediated genetic effects of biological relatives (Eaves, Pourcain, Smith, York, & Evans, 2014; Fletcher, 2011; Fletcher & Lehrer, 2011; Kong et al., 2018; Young, Benonisdottir, Przeworski, & Kong, 2019; Ystrom & Eilertsen, 2019). (See Cheesman et al. (2020) and Hwang, Davies, Warrington, and Evans (2020) for more discussion on this.) **2. Evocative rGE** occurs when an individual’s genetically

influenced traits evoke a particular response from the environment, leading to a correlation between genotype and environment. For example, a child may be genetically predisposed towards characteristics that evoke negative responses from other people. In this manner evocative rGE can lead to measures of parental negativity correlating with the child's genotype, and thus being heritable (e.g. Neiderhiser, Reiss, Lichtenstein, Spotts, & Ganiban, 2007; Neiderhiser et al., 2004). **3. Active rGE** occurs when an individual seeks out an environment that matches their genotype. For example, an adolescent may actively seek a peer group composed of adolescents similar to them. In this manner active rGE could lead to a measure of peer group deviancy being heritable (Kendler et al., 2007).

Crucially, any of these forms of rGE can lead to an association between a putative environmental exposure and an outcome without the environmental exposure causally influencing the outcome. Similarly, when an environmental exposure does have an effect, rGE may inflate the magnitude of the association between environment and outcome. To extend the 3 examples above: **1. Passive rGE** can inflate associations between parent and child traits. For example, exposure to maternal depression during childhood is predictive of child emotional and behavioural problems, but this association is inflated by genetic overlap between parent and child traits (Gjerde et al., 2019). **2. Evocative rGE** can lead to associations between adolescent behavioural problems and parental negativity, even if parental negativity does not necessarily cause adolescent behavioural problems (McAdams, Gregory, & Eley, 2013). **3. Active rGE** may inflate the association between peer group deviancy and adolescent delinquency (Kendler, Jacobson, Myers, & Eaves, 2008).

Clearly then, accounting for rGE is important where researchers want to know the extent to which a putative environmental risk factor may have a direct effect on an outcome of interest. That said, it is worth noting that while rGE is often conceptualised and treated as a cause of confounding between exposure and outcome, there are situations in which rGE can provide a tool through which the effects of environmental exposures can be assessed. For example, where variants are causally predictive of exposure they can be used as instrumental variables to assess links between exposure and outcome (Gage, Smith, Ware, Flint, & Munafò, 2016; Millwood et al., 2019; Richmond & Davey Smith, 2020).

### **Multivariate Twin Models and Causal Inference**

A major challenge when testing associations between exposures and outcome in non-experimental research settings is the effective identification and measurement of confounding variables. Confounders may be unknown and when not accounted for they will have the effect of biasing estimates of associations. Furthermore, it is likely that most associations could be confounded by many shared influences. For example, the consumption of alcohol has been shown to correlate with a range of health outcomes in positive and negative directions. Most of these associations could potentially be confounded by a wide range of socioeconomic, cultural, dietary and health-behaviour related influences. Identifying and controlling for all of these confounders would be a huge (and essentially impossible) challenge, and including them all in a regression analysis would reduce the power of that analysis to identify any remaining effects. Even if all relevant confounders could be included, error in the measurement of these confounders could still vitiate the ability to statistically control for them, and thus lead to biased effect estimates. Twin data can be used to circumvent many of these issues by controlling for genetic and shared environmental influences common to exposure and

outcome. In this manner, multivariate twin models can control for common causes without the need to identify and measure specific confounders.

It is beyond the scope of this paper to review all multivariate models to date applied to twin data but suffice to say that such data can be used to explore aetiological overlap between traits, and between exposures and outcomes, in many ways. For the purpose of this review we focus on models that are (or can be) used in attempts to draw causal inference. We do not claim to cover all such models but the underlying logic that we discuss will typically extend to twin models we do not cover.

Where univariate twin models are used to decompose the variance of a trait into genetic and environmental influences, multivariate models are concerned with decomposing *covariance* between traits in a similar manner. Figure 2 shows how to specify a twin model such that the covariance between two variables, X and Y, is decomposed into A, C and E components in the manner of a Cholesky decomposition.

INSERT FIGURE 2 AND TABLE 1 AROUND HERE

The Cholesky decomposition disaggregates the variance in trait X into genetic (A1), shared-environmental (C1) and non-shared-environmental (E1) factors, which have their effects on X via paths  $a_{11}$ ,  $c_{11}$ , and  $e_{11}$ . These factors are also able to account for variance in trait Y via paths  $a_{21}$ ,  $c_{21}$ , and  $e_{21}$ . It is these paths that account for the covariance between trait X and Y. By comparing MZ and DZ cross-twin cross-trait correlations (e.g. correlations between trait X in twin 1 and trait Y in twin 2) it is possible to decompose the covariance between traits. The residual variance in trait Y (that not accounted for by A1, C1, and E1) is decomposed into A2, C2 and E2 factors. In Table 1 we illustrate this with a variance-covariance matrix that details how the Cholesky decomposition is specified.

In Figure 3 we demonstrate how the decomposition of covariance between traits can be used to assess whether an association remains and/or is attenuated after accounting for the effects of genetic and shared-environmental influences common to both traits. Figure 3.i shows a standard linear regression of Y (the outcome or dependent variable) on X (the exposure or independent variable). This model assumes a causal relationship running from X to Y via  $B_{yx}$ . However, this relationship may be confounded by a common causal variable. In Figure 3.ii we show how such a potential confounding variable may operate, and how to model this possibility, wherein both X and Y are regressed on a potential confounding variable, Z. In this model  $B_{yx}$  is the effect of X on Y after controlling for the confounding effects of Z<sup>3</sup>. In Figure 3.iii we show the partial path diagram for the previously introduced bivariate Cholesky decomposition (for a single individual. In Figure 2 we showed how the same model is specified for twin pairs). We have rearranged the model slightly to highlight similarities to the model in Figure 3.ii and to demonstrate how it can be used to control for latent genetic and environmental confounders without the need to identify and measure them. Here, the covariance structure of the data combined with the logic of the twin model enables the decomposition of covariance into A, C and E components. In this model,  $A_1$  correlates between members of a twin pair (1.00 for MZ twins and .50 for DZs) and indexes genetic influences on X and on correlations between X and Y (within and between twins, as specified in Figure 2 and Table 1). In this manner  $A_1$  accounts for the confounding effects

---

<sup>3</sup> Note that simply regressing Y on both X and Z would have the same effect –  $B_{yx}$  would be corrected for the confounding effects of Z. This is the approach most often taken in a linear regression framework.

of common genetic influences affecting both X (the exposure) and Y (the outcome).  $C_1$  is perfectly correlated between twins of both zygosity and indexes shared environmental influences on X and on correlations between X and Y (within and between twins). As such  $C_1$  accounts for familial environmental influences that may confound the association between X and Y.  $E_1$  does not correlate between twins so only indexes the within-person correlation between X and Y after accounting for the genetic and environmental influences that make members of a twin pair similar to one another. In this manner  $A_1$  and  $C_1$  account for latent genetic and shared environmental confounders in the same way that Z accounts for measured confounders. The advantage in using twin data is that researchers do not have to identify all of the potential confounders that might account for or inflate the association between X and Y. By using twin data, all potential confounders shared by members of a twin pair (genetic and shared-environmental influences) can be accounted for.

INSERT FIGURE 3 AROUND HERE

In Figure 3.iv we show how twin data can be used to create a model very close to that of 3.ii. Arguably, given that  $B_{yx}$  is not recovered from a Cholesky decomposition, the model in 3.iv provides the “cleanest” test when using twin data to assess whether the association between X and Y survives correction for common confounders, although Cholesky decompositions are more often applied in the literature. Importantly though, the models shown in Figures 3.iii and 3.iv will result in the same conclusions being drawn if fitted to the same data. That is, if  $B_{yx}$  explains covariance between X and Y in model 3.iv then the path running  $e_{11} * e_{12}$  in model 3.iii will also do so. Either of these paths can be used to evaluate whether the association between X and Y persists after accounting for common genetic and shared-environmental influences (Turkheimer & Harden, 2014). It should be noted however that such a residual association does not on its own constitute evidence of causality for at least 2 reasons: First, non-shared environmental confounders are unaccounted for and could still explain some or all of the association between X and Y. It is of course possible to further control for additional observed non-shared environmental confounders (e.g. differences in birthweight or baseline levels of the outcome). However, challenges encountered in identifying and measuring specific non-shared environmental influences may make this a challenging task (Neiderhiser, Reiss, & Hetherington, 2007). Second, reverse causation can bias estimates resulting from these models, i.e. the estimate of the effect of X on Y can (partially) reflect the effect of Y on X. Longitudinal models with clear temporality (e.g. birth weight to ADHD symptoms) or repeated measures of the exposure and the outcome can mitigate but not always fully account for reverse causation.

### **The co-twin control design**

The application of structural equation models to decompose covariance between variables is one technique through which researchers can use multivariate twin data to ask the question “does the association between X and Y persist after accounting for genetic and shared-environmental confounds?” Another method involves strengthening causal inference by using one twin in a pair as a direct matched control for the other. This approach is known as the 'co-twin control design', in the field of behavioural genetics and as one implementation of the 'fixed effects' approach in econometrics. While the biometrical decomposition of variance discussed above is most often used to dissect the

aetiological architecture of phenotypes, the main purpose of the co-twin design is causal inference.

The co-twin control design is best understood within the counterfactual framework for causal inference (Lynch, 2020). In the (theoretical) counterfactual situation, a given individual receives a treatment (or is exposed to a risk factor value) and, at the same time, the same individual does not receive the treatment. In this ideal situation, the treated individual is literally the same as the control individual, the only difference being the treatment. It thus becomes possible to estimate the causal effect of that treatment, even at the individual level. Naturally, in real life, the same individual cannot at the same time receive and not receive a treatment, so strict causal inference is impossible. However, all causal inference designs can be conceived as aiming to approximate this ideal counterfactual situation (Pingault et al., 2018). One way to approximate this situation is to find the best match possible. This is when twins, and in particular MZ twins, come into play. Indeed, an exposed twin shares all of their genetic material and, by definition, their shared environment, with the non-exposed co-twin. At a basic level, the co-twin design involves comparing outcomes between the exposed and non-exposed twin, and thus controlling for genetic and shared environmental influences. If the risk factor (e.g. smoking) causes the outcome (e.g. lung cancer) we should expect that on average smoking twins will have a higher incidence of lung cancer than their non-smoking co-twins. If there was no difference in the incidence of lung cancer in the smoking vs non-smoking twin, then this would suggest the relationship between the risk factor and the outcome is non-causal and due to genetic or shared-environmental confounding.

The co-twin control design is often referred to as the 'discordant twin design' when the exposure is binary, and the 'twin differences design' when the exposure is continuous. A comprehensive review of regression models used to fit the co-twin design can be found in Carlin, Gurrin, Sterne, Morley, and Dwyer (2005). For continuous exposure and outcome, twin differences in the exposure and the outcome are first computed. The causal effect is estimated by a simple regression without intercept of the twin differences in the outcome on the twin differences in the exposure. A more complex multilevel model can be fitted, modelling both the within family-estimates and the between-family estimates. The within-family estimates will give exactly the same estimates of the causal effect as the aforementioned simpler approach. This type of model must allow for both an intra-class correlation (the within-twins correlation), which affects the estimate, and robust standard errors. The phenotypic relationship between the exposure and the outcome is a weighted mean of the within and between family estimates. They can be fitted in a variety of statistical frameworks which provide very similar estimates and confidence intervals, including mixed models, generalized estimating equations, and structural equation modelling.

The co-twin control design has been widely used for causal inference, confirming for example the causal link between smoking and lung cancer (Hjelmberg et al., 2017), demonstrating the impact of bullying victimisation on childhood mental health (Silberg et al., 2016), and suggesting that birth weight impacts the development of ADHD symptoms (Lim et al., 2018). Findings from the co-twin design can shed new light on the role of risk factors. For example, simple associations suggest that bullying victimisation is a risk factor for most mental health outcomes over the long term, whereas twin differences designs point towards more specific effects (e.g. prominently on anxiety, depressive symptoms and self-harm) and shorter-term effects (Singham et al., 2017). More elaborate longitudinal models can also be built based on twin differences. For

example, a latent growth model has been implemented to show that differences in birth weight predict ADHD symptoms from childhood to adolescence, but that the size of the effect decreases with age, consistent with a partial 'catch-up' effect (Lim et al., 2018). Cross-lag models can be implemented with twin differences to look at reciprocal influences between two phenotypes over time, while controlling for genetic and shared environmental confounding (Cecil, Barker, Jaffee, & Viding, 2012).

#### *Comparing the co-twin control design with the classical twin decomposition*

The co-twin control design and the classical twin decomposition can be conceived as two approaches to decomposing a relationship: The former decomposes covariance between variables into within and between family effects, the latter into genetic, shared-environmental and non-shared-environmental components. In the majority of instances, when using twin samples for causal inference, these two methods are equivalent and should lead to the same conclusions. That is, when using the co-twin control design to focus on differences between MZ twins (within family effects), researchers are in fact focussing on non-shared environmental variance (the E component in biometric twin models). When asking if differences in X predict differences in Y, they are asking if non-shared environmental variance in X predicts non-shared environmental variance in Y. Such statistical equivalence does not mean that one or the other method should be considered redundant. Twin decompositions will be of use to researchers interested in quantifying the relative effects of genetic, shared environmental and non-shared environmental correlations on covariance between variables. Conversely, co-twin control studies are not appropriate to examine the genetic architecture of traits; they have been used and are most useful for causal inference research, as a particular implementation of fixed-effect models.

In the context of causal inference, both biometric decomposition and twin differences approaches suffer from the same two major limitations of unobserved non-shared environmental confounders and the possibility of reverse causation. However, while twin differences and biometric decompositions are usually equivalent, there may be circumstances in which one design is appropriate and the other is not. For example, a version of the twin difference design has been applied to a sample of genotyped DZ twins to explore the extent to which differences in polygenic scores predicted differences in several early life outcomes (Selzam et al., 2018; Selzam et al., 2019). The intention was to control for the potential confounding effects of those influences shared within DZ twin pairs including population stratification, assortative mating and indirect genetic effects from parents (genetic nurture). Genetic nurture occurs for example when genetic variants for education in the parents affect parental education, which in turn environmentally impacts child education. Such effects, although originating in the parental genotypes, are environmentally mediated and do not reflect direct genetic effects of the child genotype on the child phenotype (Eaves et al., 2014; Hwang et al., 2020; Kong et al., 2018; Young et al., 2018). Selzam et al. (2019) showed that twin differences in polygenic scores predicted differences in height, BMI, intelligence, educational achievement and ADHD symptoms. However, the within-family prediction was attenuated for intelligence and educational attainment, indicating a degree of confounding. This finding required the use of a twin differences approach. A biometric decomposition would not be appropriate here as it would be inappropriate to treat polygenic scores as phenotypes (as any polygenic score would correlate 1 within MZ twins and 0.50 on average in DZ twins, resulting in  $A = 100\%$ , which would not capture and adjust for the genetic nurture effects).

## Direction of Causation in Twin Models

### *Longitudinal data*

In the models we have discussed thus far the assumption has been that the direction of causation runs from X to Y and not vice versa. However, often it is not possible to know with certainty whether this is the case. For example, in cross-sectional data it may well be that Y causes X but it would not be possible using the models we have discussed to identify the true direction of causation. An often-used technique to identify the direction of effects is to use longitudinal data in which variables have a temporal order to them. For example, in a bivariate Cholesky decomposition, if X was measured before Y, one could argue that the data should be modelled such that X predicts Y and not vice versa. That said, temporal ordering of measurement is often not justification enough to assume that a variable at time 1 should be modelled as a cause of a variable at time 2. For example, if school performance were measured at time 1 and IQ at time 2, the temporal ordering of measurement would not justify modelling the data based on the assumption that school performance *causes* IQ. Ideally, IQ and school performance would each be measured at time 1 *and* time 2, so that we can ask whether school performance prospectively predicts IQ after accounting for the correlation between IQ and school performance at time 1 and stability in IQ between time 1 and time 2.

Autoregressive cross-lagged models are frequently used so that prospective associations between variables can control for the combined effects of contemporaneous associations between traits and across-time stability within traits. For example, in Figure 4i, the prediction from X1 to Y2 controls for the effects of Y1 on Y2 and the covariance between X1 and Y1 (although see Hamaker, Kuiper, and Grasman (2015) for a critique of some misperceptions regarding autoregressive cross-lagged models and alternative specifications). A bivariate Cholesky decomposition of X1 predicting Y2 would potentially capture the confounding effects of Y1 through the genetic and shared-environmental influences common to all, but would not capture potential non-shared environmental confounders indexed by Y1.

Malanchini et al. (2017) introduce a biometric genetic version of the autoregressive cross-lagged model in which four variables are decomposed into genetic, shared environmental and non-shared environmental components, and each of these components allowed to predict one another in the manner of an autoregressive cross-lagged simplex model (see Figure 4ii-4iv). In this model, the cross-lagged E predictions tell us whether e.g. X1 predicts Y2 after accounting for genetic and shared-environmental influences common to these two variables *and* those influences—including non-shared environmental influences—shared with Y1<sup>4</sup>. Malanchini et al. (2017) demonstrated the utility of this model by showing that although the phenotypic cross-lags between reading motivation and reading achievement in children were of equal magnitude, only the prediction from motivation to achievement survived correction for shared aetiologies.

INSERT FIGURE 4 AROUND HERE

An alternative biometric autoregressive cross-lagged model to that presented here comprises a combination of the Cholesky decomposition and the correlated factors

---

<sup>4</sup> Note that this model is distinct to previous autoregressive cross-lagged twin models such as those critiqued by Luo, Haworth, and Plomin (2010) in which correlations between X1 and Y1 and between the residual variances of X2 and Y2 are decomposed into A, C and E components, but cross-lagged paths are not decomposed.

model (where e.g. in Figure 4.ii causal effects would run from Ax1 and Ay1 directly to X2 and Y2 rather than via Ax2 and Ay2. See: Torvik et al., 2019; Torvik et al., 2017). This model has been used to show that social anxiety disorder prospectively predicts alcohol use disorder—and not vice versa—after accounting for shared aetiology (Torvik et al., 2019).

#### *Cross-sectional data and Direction of Causation (DoC) twin models*

While longitudinal data is typically considered ideal for identifying the direction of effects, it isn't always available. In such cases cross-sectional twin data can, to some degree, help strengthen causal inference between two traits X and Y. The Direction of Causation (DoC) model (Duffy & Martin, 1994; Heath et al., 1993) is distinct to most of the models described here in that the focus is not on controlling for the confounding effects of common aetiologies. Rather it is on determining the likely direction of effects between two variables where temporal ordering is not present or is unknown (see Figure 5.i). That is, the extent to which an association between X and Y is driven by X predicting Y vs. Y predicting X. The power of this design is derived from the differential predictions that come about regarding the cross-twin cross-trait correlations for MZ and DZ twin pairs when comparing the two potential causal models (X causes Y, vs Y causes X). Differences in the covariance structure are predicted where the aetiological structure (i.e. the relative importance of A, C, and E) is distinct for trait X vs. trait Y. For example, where A affects trait Y but not X and C affects trait X but not Y.

INSERT FIGURE 5 AROUND HERE

Figure 5 contains examples in which trait X is under the influence of C and E, and trait Y is under the influence of A and E. In 5.ii, trait X causes Y, in 5.iii, Y causes X. If X causes Y, the expected cross-twin cross-trait correlation (i.e., correlation between X in twin 1 and Y in twin 2, or Y in twin 1 and X in twin 2), can only be explained via the variance of variable X. Since variable X is influenced by C and E only, the cross-twin cross-trait correlations will be driven by C. If C is the only thing that explains correlations between e.g. X in twin 1 and Y in twin 2, then these correlations will be the same for MZs and DZs (because C, the shared environment, by definition influences MZ and DZ twins equally). However, if Y causes X, then the cross-twin cross-trait correlations will be driven by variance in variable Y. Given that Y in our example is under the influence of A, the cross-twin cross-trait correlations will be larger for MZs than DZs (Gillespie & Martin, 2005). While our example involves traits of very different aetiologies, it is also possible to distinguish the direction of causation when the two traits have a similar but distinct aetiology. For example, where trait X is under greater genetic influence than trait Y. In this manner it has been shown that for the relationship between autism spectrum disorder and other mental health problems, best-fitting models suggest a unidirectional phenotypic influence of hyperactivity on identification of ASD. Relationships between ASD and emotional symptoms and conduct problems were best explained by correlated genetic and environmental factors, respectively (Tick et al., 2016).

A major assumption of the DoC model is that the association between X and Y is causal. That is, that it is not attributable to common causes unmodelled. In theory it is possible to allow for correlations between aetiological components of X and Y (i.e. between A<sub>1</sub> and A<sub>2</sub> or C<sub>1</sub> and C<sub>2</sub>), however doing so will reduce the ability of the model to distinguish the direction of effects. In practice it is more usual for researchers to compare e.g. a Cholesky decomposition of Y on X in which a shared aetiology is modelled, with a

DoC model and then select the model with the best fit to the data (e.g. Gillespie et al., 2012).

One way of thinking about the DoC model is that the variance components of one trait are being treated as instrumental variables in their prediction of the other trait. That is, that their effects only operate through the causal effect of the intermediate trait. For example, that  $A_2$  can only predict Trait X through Trait Y. This is an assumption of the model and is analogous to the use of genetic variants and polygenic scores as instrumental variables in those methods collectively referred to under the umbrella term Mendelian randomization (Davey Smith & Ebrahim, 2003; Davey Smith & Hemani, 2014; Smith & Ebrahim, 2005) Hwang et al. (2020) discuss extensions to the DoC model that incorporate Mendelian genetic tools to further strengthen the ability to draw causal inferences using this model. An alternative version of the DoC has also been suggested, in which the direction of causation between latent C and E factors (as opposed to between phenotypes) can be detected using datasets with non-normally distributed variables (Ozaki & Ando, 2009).

### **Extended Family Designs**

Twin models make use of the fact that MZ and DZ twins differ in how genetically related they are to one another, and this information is used to infer how important genetic and environmental factors are in explaining variance in a trait and covariance between traits. However, the logic of the twin design is not confined to use in twin samples only. Indeed, it is possible to apply biometric ACE models to any datasets comprising differentially related relatives. For example, it is possible to use samples comprising siblings and half-siblings, or cousins and half-cousins to estimate heritability and to explore the nature of covariance between phenotypes<sup>5</sup>. Various extended family models have been proposed that make use of multi-generation data to examine the aetiology of traits measured in two or more generations (Keller et al., 2009; Maes et al., 2009; Truett et al., 1994). Such models allow for the modelling of assortative mating, gene-environment covariance, and other possibilities not feasible in a single-generation dataset. Extended family data also provides the opportunity for researchers to examine associations *between* family members. That is, to ask questions about whether family members have an effect on one another, after accounting for their relatedness. For example, using data comprising twins and their children, it is possible to assess whether associations between parents and children persist after accounting for genetic relatedness (D'Onofrio et al., 2003; McAdams et al., 2014; Silberg & Eaves, 2004).

#### *Children-of-Twins-and-Siblings Models*

Often, researchers are interested in the potential for parents to influence the development of their children. For example, associations between parenting behaviours and child behavioural and emotional traits are often interpreted in ways that imply that parenting *causes* child behaviour (and/or vice versa). And intergenerational similarities in personality, psychopathology, and behaviour are often interpreted as though exposure to parent trait causes the development of the child trait. However, in most instances, parents and children are genetically related to one another and share an environment. As

---

<sup>5</sup> There are limitations to using such samples. For example, the lower relatedness coefficients will mean that statistical power is reduced relative to similarly sized samples of twin pairs. Where twins are always age matched, siblings and cousins may not be, so appropriate controls/adjustments may need to be included in models where necessary. Assumptions regarding the definition of the 'shared environment(s)' will also need to be made explicit.

such parent-child associations are likely to be at least partially attributable to the confounding effects of shared genetic and environmental influences.

Using samples comprising twins and their children, it is possible to decompose intergenerational associations into genetic and environmental components in much the same way that samples of twins can be used to decompose within-person associations. Where twin studies use differences in MZ and DZ cross-twin cross-trait correlations, children-of-twin studies use differences in avuncular correlations (that between a twin and their co-twin's child/ren). As with twin data, children-of-twin data can be analysed using a variety of models (Heath, Kendler, Eaves, & Markell, 1985) and can be extended to include siblings (as well as twins) with children (Hannigan et al., 2018; Kuja-Halkola, D'Onofrio, Larsson, & Lichtenstein, 2014), multiple offspring per parent (McAdams et al., 2018), and spouses (Silberg, Maes, & Eaves, 2010). In Figure 6 we present a simple version of a children-of-twins model comprising MZ and DZ twin pairs with one child per twin. In this model, both parent and child phenotypes are decomposed into genetic and environmental influences. In the child generation this is possible by comparing cousins whose parents are MZ twins to those whose parents are DZ twins ( $r_G=.25$  and  $.125$  respectively). If multiple children per twin are included, then siblings can be included in these estimations ( $r_G=.50$ ). Because cousins do not share a nuclear family environment,  $C$  (as typically defined) cannot be estimated in the offspring generation without the use of siblings. The intergenerational association between parent and child is decomposed into a genetic component ( $A1'$ ), an extended family component ( $C1'$ )<sup>6</sup> and the remaining residual phenotypic association ( $p$ ). If this ' $p$ ' path predicts the child trait, then this indicates the association between parent and child phenotypes remains after accounting for shared genetic and environmental influences.

INSERT FIGURE 6 AROUND HERE

Models similar to that illustrated in Figure 6 have been used to examine a range of associations between parent and child traits, demonstrating for example that maternal depression during pregnancy does not predict later childhood emotional and behavioural problems after accounting for genetic confounding (Hannigan et al., 2018), and that maternal smoking during pregnancy predicts birthweight and preterm birth but not cognitive abilities or externalising behaviours in offspring after accounting for genetic confounds (Kuja-Halkola et al., 2014). While most of these analyses are bivariate, these models are just as flexible as other twin models, so where appropriate data is available it will be possible for example to ask questions about the direction of causation between parent and child (Narusyte et al., 2008), longitudinal associations between parent and child, and gene-environment interplay.

### **Twin Data and Experimental Design**

Thus far we have discussed *quasi-experimental* research designs using samples of twin pairs. However, another (rarely employed) approach involves designing *experiments* using twin samples. Typically, experiments involve the random assignment of individuals to treatment and control conditions, with this random assignment intended to account for any confounders that might exist in a non-experimental situation. Matched-pair designs go one step further and match pairs of subjects in the control and experimental

---

<sup>6</sup> In extended family data it is possible to model a range of different "shared environments". For example, maternal effects can be modelled in studies including maternal and paternal half siblings.

conditions on any potentially important factors that might make pairs differ from one another. Using twin data such that twin 1 receives a treatment and twin 2 does not provides an extreme example of this matched-pair design. Where twin 1 and 2 are members of an MZ pair they will be matched on all genetic and shared environmental influences. Examples of this method include studies demonstrating how supplementation with vitamin C has no effect on self-reported incidence of the common cold (Carr, Einstein, Lai, Martin, & Starmer, 1981; Martin, Carr, Oakeshott, & Clark, 1982).

An alternative take on the use of twins in experiments comes from a study by Haworth et al. (2016), in which they estimated changes in aetiological influences on wellbeing and mental health in response to an online intervention delivered to all twins in a sample. They found that mean levels of wellbeing and mental health improved across the study, and that changes in wellbeing were accounted for by changes in non-shared environmental influences. However, while means changed and new non-shared environmental influences were evident post-intervention, estimates of the relative importance of A, C and E in explaining variance did not vary in magnitude pre- vs. post-intervention. These findings highlight a fundamental characteristic of biometric (ACE) twin models: such models focus on population variance in traits, not population means. Means can change without necessarily affecting the relative importance of genetic, shared environmental and non-shared-environmental components.

## **Discussion**

In the preceding review we have covered several ways in which twin data can be used by researchers interested in testing quasi-causal and causal hypotheses. As with all statistical models, those we have presented require careful and considered interpretation. In the following section we explore some issues that we believe are important when attempting causal inference with twin data.

### *Non-shared-environmental confounds are not accounted for*

Most of the models presented in this chapter are focussed on controlling for genetic and shared-environmental sources of confounding, and do not do anything to control for potential non-shared confounders. As such, the potential for non-shared environmental influences to confound associations should always be considered. For example, in an analysis in which smoking is the predictor, twins who smoke more than their co-twins may also drink more alcohol, such that alcohol rather than the putative risk factor - smoking- may be responsible for adverse outcomes. It is possible to statistically control for twin differences in covariates—e.g. alcohol consumption—but the models presented will still assume that there are no additional (unobserved) non-shared environmental confounders. To the extent that there are non-shared environmental confounders that are not accounted for, any residual association between predictor and outcome will be inflated.

### *Some environments of interest do not differ between twins*

It is worth noting that some environmental variables, such as childhood socioeconomic status, or neighbourhood characteristics during childhood, will not typically vary between twins. That is, where twins are reared by the same parents, in the same household, go to the same school and live in the same neighbourhood, then they may not differ in terms of environmental measures related to parents, households, schools and neighbourhoods. This precludes the estimation of heritability of such variables using the twin design (i.e. MZ and DZ correlations will both be 1 in these cases). It also precludes

the calculation of difference scores in the co-twin control design. Importantly, this does not necessarily mean that such variables are not under the influence of genetic factors, it is simply a limitation of the twin design that heritability cannot be estimated for variables that do not differ within twin pairs. As such, it is not possible to correct for aetiological overlap involving such variables.

One solution to this problem lies in the use of intergenerational multiple-children-of-twins-and-siblings models. If a dataset comprises twins and siblings in the parent generation as well as the offspring generation (which would also then include cousins), it becomes possible to model the aetiology of traits that can differ between siblings in the parent generation, such as e.g. parental socioeconomic status. By modelling genetic and environmental relatedness between parents and children it then becomes possible to account for genetic overlap between parent trait and child outcome. For example, this method has been used to demonstrate that parental educational attainment is predictive of child ADHD symptoms and academic problems at school, but not child depression symptoms, after accounting for genetic relatedness (Torvik et al., 2020).

### *Statistical power*

Much has been written about statistical power and twin studies (e.g. Ahmadzadeh et al., 2020; Verhulst, 2017; Visscher, 2004). In univariate studies, power to detect A (defined as twice the difference between MZ and DZ correlations) is contingent upon MZ and DZ correlations being significantly different to one another. Power to detect C is typically lower than that to detect A, and relies on DZ twins being more similar to one another than heritability accounts for. E will always be non-zero because it contains error variance.

When using bivariate and multivariate models as discussed in this chapter, the focus is on whether residual associations remain after accounting for potential genetic and shared-environmental confounds. While E (residual variance after accounting for A and C) will always be non-zero in univariate models that explain variance in a single trait, this is not the case for covariance between traits, where residual covariance attributable to E can equal zero. As such, when decomposing covariance between traits for the purposes of causal inference there should be 2 primary concerns relating to power:

1. How much power is there to detect genetic and shared-environmental influences on the covariance?
2. If power to detect genetic and shared-environmental influences on covariance is adequate, how large would the residual covariance need to be to be detected?

Where covariance between traits is low, or sample sizes are small, power to detect A or C influences on covariance will be low, and this may lead to incorrect conclusions being drawn regarding the significance and/or magnitude of residual covariance terms. Researchers should therefore be careful when interpreting residual covariance and consider how much power they have to detect confounding. Power calculations can be performed by simulating a dataset in which the sample composition, covariance, and proportions of covariance attributable to A, C and the residual/E are known. By varying these proportions researchers can evaluate the power they have in their own data (e.g. see Verhulst, 2017, and associated website).

### *Measurement error*

Measurement error is aggravated when computing difference scores, to the extent that traits are correlated within pairs. As a result, estimates from MZ differences may be attenuated compared to estimates in DZ differences, which themselves are attenuated compared to phenotypic estimates, even in the absence of confounding. This may lead to

an underestimation of causal effects. Causal estimates can also be biased when the magnitude of measurement error is different for the risk factor and the outcome. This can be dealt with by directly modelling measurement error when possible, or by conducting a sensitivity analysis to estimate how much measurement error would be needed to substantially change the conclusions.

### *The importance of context*

When using twin data to account for genetic or environmental overlap between two traits, one can think of the estimate of shared genetic influences as a 'bivariate heritability' estimate – an estimate of genetic influence on covariance (rather than variance). Like heritability estimates, these bivariate heritability estimates are population specific and should not be assumed to necessarily generalise to other populations. Estimates of genetic and environmental influences on variance are known to vary with age (Haworth et al., 2010), across socioeconomic gradients (Tuvblad, Grann, & Lichtenstein, 2006), and between countries (Ball et al., 2009; Hur, 2008; Xu et al., 2015). In other words, contextual conditions have an effect on how much variance in a trait is explained by genetic and environmental variance in a population. This is also likely to be true for bivariate estimates and thus has implications for the use of twin data for the purposes of causal inference: if we find that the association between X and Y is entirely attributable to genetic overlap in our sample, this does not necessarily mean that the same will be true in other samples. Indeed, it may even be possible that the nature of the association varies *within* a sample. When this is deemed plausible, it is worth considering the inclusion of moderators in models where feasible. For example, it is conceivable that parental socioeconomic status (SES) could moderate the nature of the association between maternal depression and child internalising such that at high levels of SES (where social and economic resources are abundant), the association may be entirely genetic, but at low levels of SES (where support may be less available), there could be a causal effect of maternal depression on child internalising. Where moderation is taking place but is not modelled, estimates regarding the association between X on Y will be biased.

### **Summary**

In this chapter we have described some of the ways in which twin data can be used to interrogate causal hypotheses, particularly where experimental manipulation is not possible. The advantage of using twin and family data lies in the ability to control for latent genetic and shared environmental sources of potential confounding. This provides researchers with a powerful alternative to the often prohibitively challenging task of explicitly identifying and collecting data on potential confounds. We would argue that where an association of interest survives correction for the potential confounding effects of common genetic and shared-environmental effects, then this should strengthen the argument for drawing causal inferences from the data (always bearing in mind the quasi-experimental nature of the design used). For example, if X predicts Y after controlling for latent genetic and shared-environmental effects that influence both variables, then it is possible that X causes Y and less likely that the association is attributable to common causes. If so, then interventions aimed at increasing or reducing X may have an effect on Y. However, if the association between X and Y does not persist after controlling for latent genetic and shared-environmental confounds, then it is unlikely that X causes Y. It should therefore be considered less likely that intervening to alter X will alter Y also.

In discussing the potential inherent in using twin and family data to interrogate causal hypotheses, we have highlighted some of the limitations, assumptions, and nuances that researchers should consider when doing so. It is worth noting however that many of these limitations are shared with other research methods and broadly relate to the generalisability of findings, and the importance of being cognisant of assumptions when interpreting results. Providing these issues are borne in mind, twin studies provide an extremely powerful tool for researchers interested in teasing apart potential causality from the effects of unknown confounders – a fundamental and central goal of all scientists concerned with the study of human beings.

For readers interested in using twin data for causal inference, a list detailing twin registries around the world, with links to detailed overviews of each, is published periodically in the journal *Twin Research and Human Genetics* (see Hur et al. (2020) for the most recent).

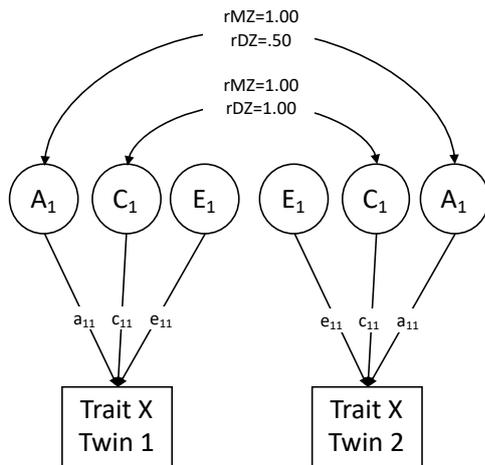


Figure 1. Path diagram of the univariate twin model.

All of our path diagrams follow standard structural equation modelling convention whereby measured variables are represented with squares and latent variables are represented with circles. Although not included in the path diagram, the variances of all latent variables are fixed to unity. a<sub>1</sub>=Additive genetic effects on phenotype; c<sub>1</sub>= shared-environmental effects on phenotype; e<sub>1</sub>=nonshared environmental effects on phenotype. A<sub>1</sub>=Additive genetic variance component; C<sub>1</sub>= shared-environmental variance component; E<sub>1</sub>=nonshared environmental variance component. In the classical twin design, the predicted variances of each trait ( $a_{11}^2 + c_{11}^2 + e_{11}^2$ ) and the predictions for the covariances ( $a_{11}^2 + c_{11}^2$  for MZ and  $.5a_{11}^2 + c_{11}^2$  for DZ pairs) are fitted to the observed variances and covariances to obtain the most likely estimates for the A, C and E effects.

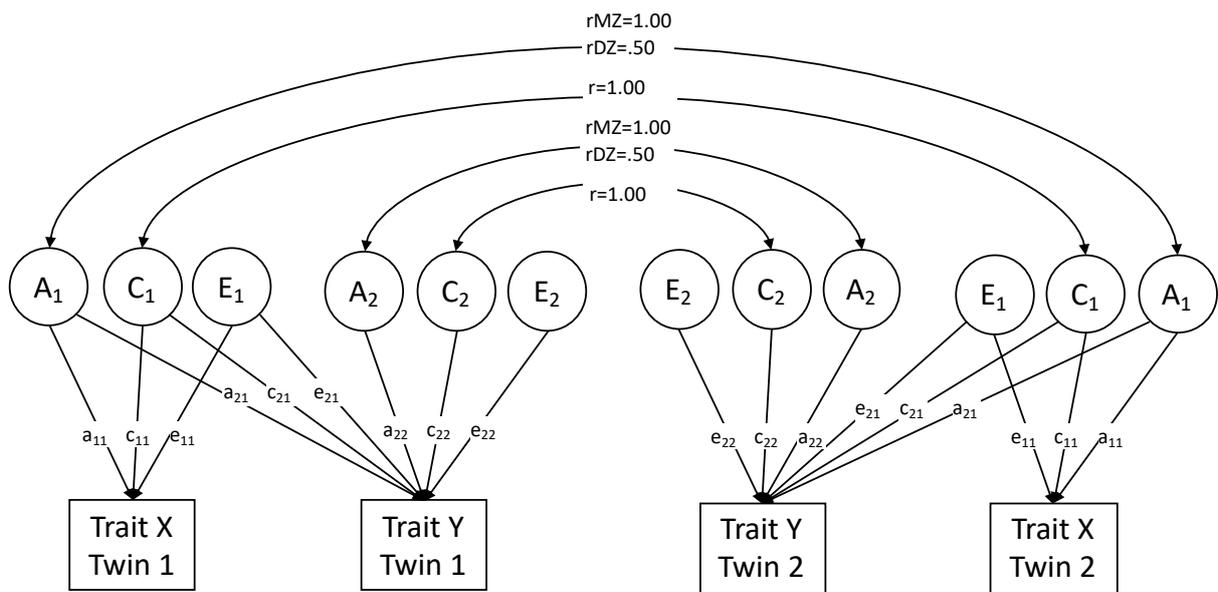


Figure 2: Cholesky decomposition twin model.

Table 1. Variance-covariance specification for a bivariate Cholesky decomposition

	Trait X Twin 1	Trait Y Twin 1	Trait X Twin 2	Trait Y Twin 2
Trait X Twin 1	$a_{11}^2 + c_{11}^2 + e_{11}^2$	$a_{11} * a_{21} + c_{11} * c_{21} + e_{11} * e_{22}$	$.5 * a_{11}^2 + c_{11}^2$	$a_{11} * .5 * a_{21} + c_{11} * c_{21}$
Trait Y Twin 1	$a_{11} * a_{21} + c_{11} * c_{21} + e_{11} * e_{22}$	$a_{21}^2 + c_{21}^2 + e_{21}^2 + a_{22}^2 + c_{22}^2 + e_{22}^2$	$a_{11} * .5 * a_{21} + c_{11} * c_{21}$	$.5 * a_{21}^2 + c_{21}^2 + .5 * a_{22}^2 + c_{22}^2$
Trait X Twin 2	$a_{11}^2 + c_{11}^2$	$a_{11} * a_{21} + c_{11} * c_{21}$	$a_{11}^2 + c_{11}^2 + e_{11}^2$	$a_{11} * a_{21} + c_{11} * c_{21} + e_{11} * e_{22}$
Trait Y Twin 2	$a_{11} * a_{21} + c_{11} * c_{21}$	$a_{21}^2 + c_{21}^2 + a_{22}^2 + c_{22}^2$	$a_{11} * a_{21} + c_{11} * c_{21} + e_{11} * e_{22}$	$a_{21}^2 + c_{21}^2 + e_{21}^2 + a_{22}^2 + c_{22}^2 + e_{22}^2$

Variations are highlighted in pale blue with black outline on the diagonal. MZ twin covariances are given below the diagonal, DZ twin covariances above the diagonal. Within-twin cross-trait covariances are highlighted in dark blue. Cross-twin within-trait covariances are highlighted in pale green. These can be used to decompose variances in X and Y. Cross-twin cross-trait covariances are highlighted in dark green. These are used to decompose covariance between X and Y. Note that e paths are never involved in covariance between twins, only within.

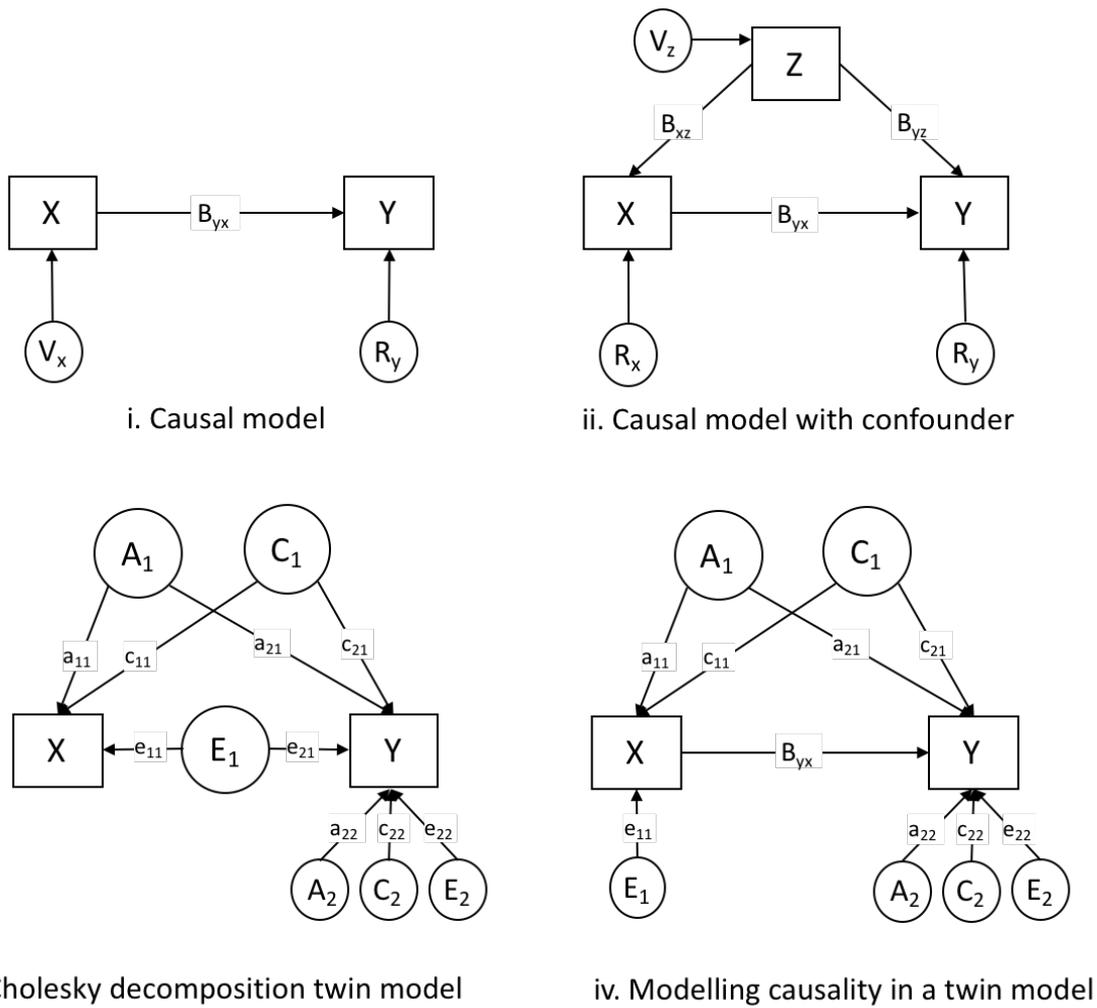


Figure 3. Modelling causality

In the above models,  $B_{yx}$  is the causal effect of X on Y. In model i.  $V_x$  denotes the variance of X, and  $R_y$  denotes the residual variance of Y (the variance of Y remaining after regressing out the effects of X). In model ii.  $V_z$  denotes the variance of Z, and  $R_x$  and  $R_y$  denote the residual variances of X and Y (variance remaining after regressing out the effect of Z (on X) and Z and X (on Y)). Twin models shown are partial path diagrams for a single individual. All latent factors have variance of 1 (not shown).

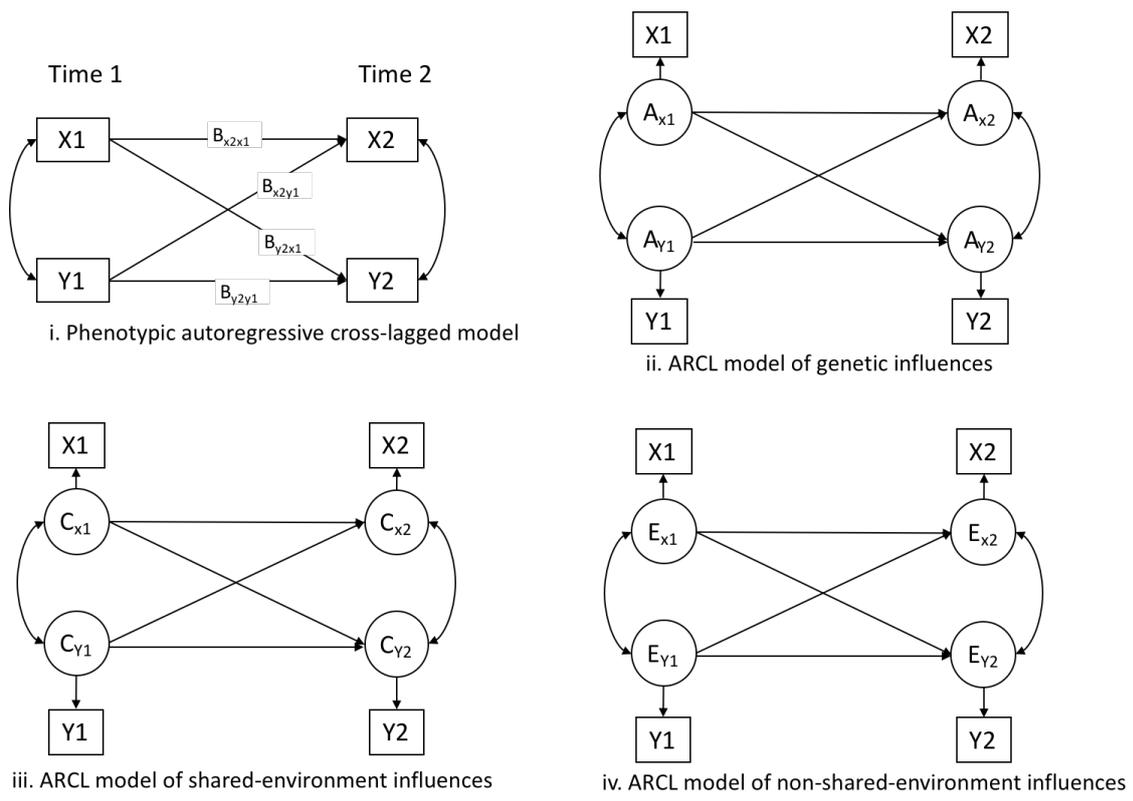


Figure 4. A phenotypic autoregressive cross-lagged model and partial path diagrams of a twin autoregressive cross-lagged model

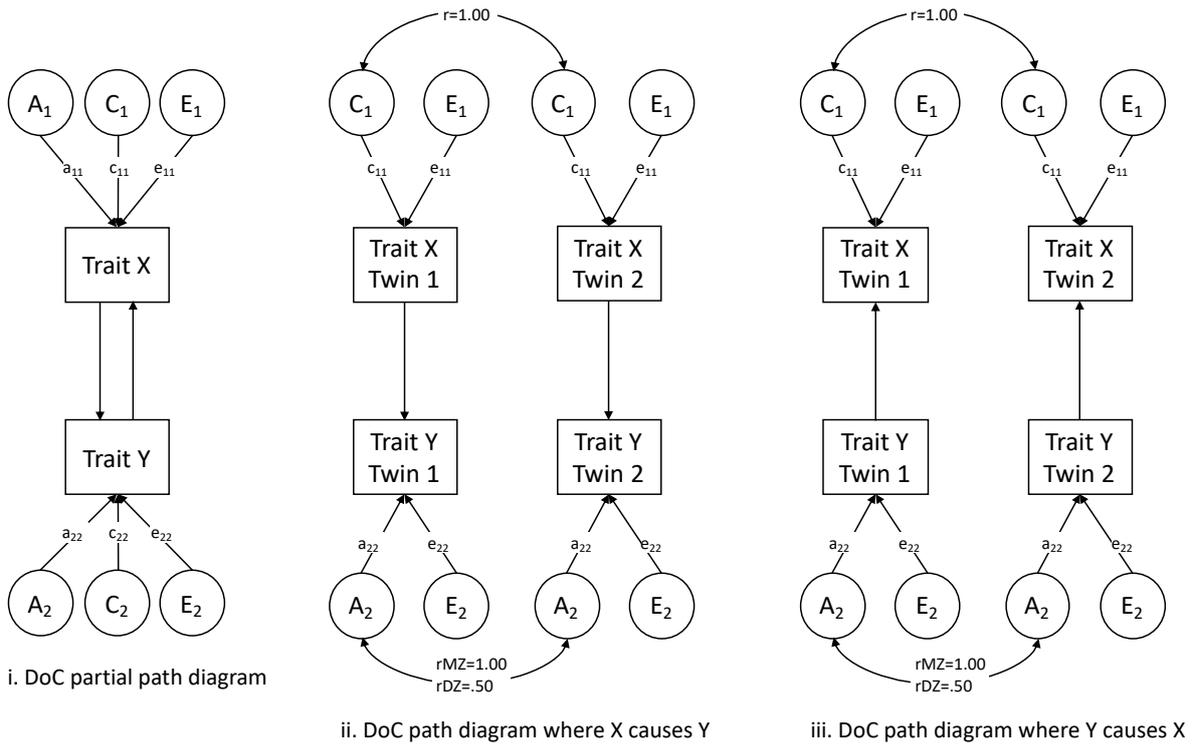
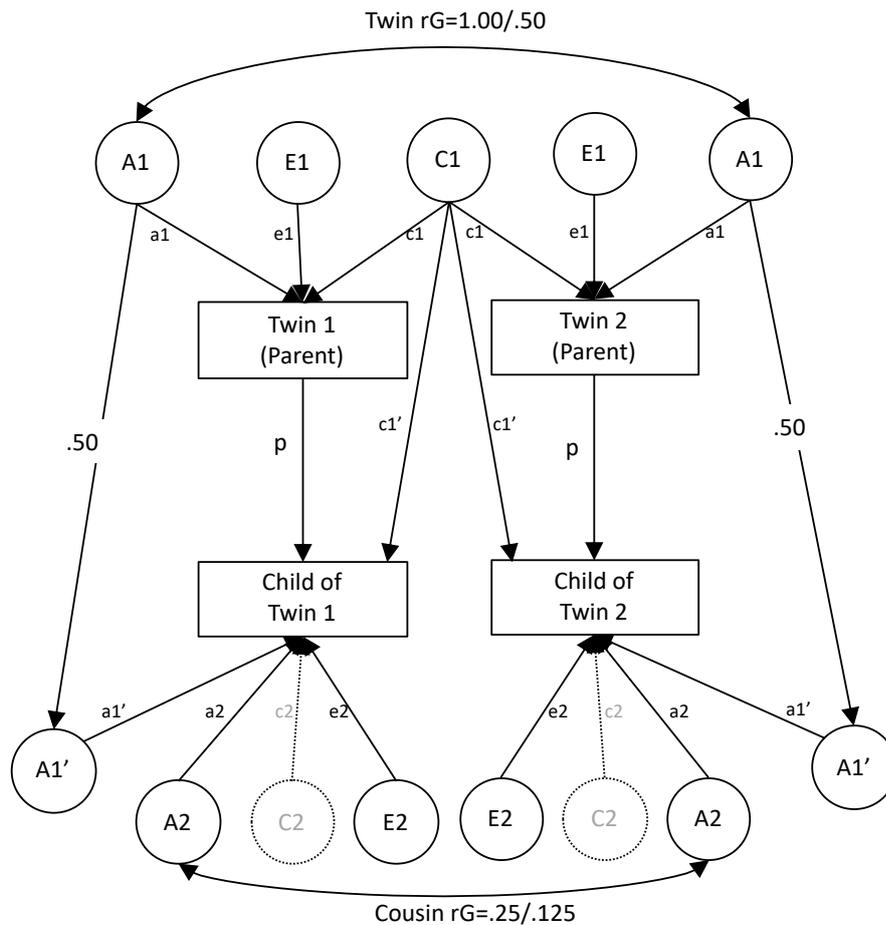


Figure 5. The Direction of Causation (DoC) twin model



**Figure 6. Children of twins structural equation model**

$a_1$ =additive genetic effects on parental phenotype;  $c_1$ = shared-environmental effects on parental phenotype;  $e_1$ =nonshared environmental effects on parental phenotype;  $a_1'$ =genetic effects common to parental phenotype and offspring phenotype;  $c_1'$  = extended family environment effects on offspring phenotype;  $a_2$ =genetic effects specific to offspring phenotype;  $c_2$ = shared-environmental effects on offspring phenotype (not estimable using cousin data);  $e_2$ =nonshared environmental effects on offspring phenotype;  $p$ =phenotypic effect of parent on offspring; NB the pathway between  $A_1$  and  $A_1'$  is fixed to .50 because parents and children share 50% of their genome. To avoid over complicating path diagrams, variance paths have been omitted, but for all latent factors variance=1. For  $A_1'$  this means that residual variance (after accounting for the path between  $A_1$  and  $A_1'$ ) is .75.

## References

- Ahmadzadeh, Y. I., Eley, T. C., Hannigan, L. J., Creswell, C., Lichtenstein, P., Reiss, D., . . . McAdams, T. A. (2020). Parental criticism and adolescent internalising symptoms: Associations remain after accounting for shared genetic effects. *medRxiv*, 2020.2005.2007.20084319. doi:10.1101/2020.05.07.20084319
- Ball, H. A., Sumathipala, A., Siribaddana, S. H., Kovas, Y., Glozier, N., McGuffin, P., & Hotopf, M. (2009). Genetic and environmental contributions to depression in Sri Lanka. *The British Journal of Psychiatry*, 195(6), 504-509.
- Carlin, J. B., Gurrin, L. C., Sterne, J. A., Morley, R., & Dwyer, T. (2005). Regression models for twin studies: a critical review. *International Journal of Epidemiology*, 34(5), 1089-1099.
- Carr, A. B., Einstein, R., Lai, L. Y., Martin, N. G., & Starmer, G. A. (1981). Vitamin C and the common cold using identical twins as controls. *Medical Journal of Australia*, 2(8), 411-412.
- Cecil, C. A., Barker, E. D., Jaffee, S. R., & Viding, E. (2012). Association between maladaptive parenting and child self-control over time: Cross-lagged study using a monozygotic twin difference design. *The British Journal of Psychiatry*, 201(4), 291-297.
- Cheesman, R., Eilertsen, E. M., Ahmadzadeh, Y. I., Gjerde, L. C., Hannigan, L. J., Havdahl, A., . . . McAdams, T. A. (2020). How important are parents in the development of child anxiety and depression? A genomic analysis of parent-offspring trios in the Norwegian Mother Father and Child Cohort Study (MoBa). *medRxiv*, 2020.2004.2014.20064782. doi:10.1101/2020.04.14.20064782
- Conley, D., Rauscher, E., Dawes, C., Magnusson, P. K., & Siegal, M. L. (2013). Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins. *Behavior Genetics*, 43(5), 415-426.
- D'Onofrio, B. M., Turkheimer, E. N., Eaves, L. J., Corey, L. A., Berg, K., Solaas, M. H., & Emery, R. E. (2003). The role of the Children of Twins design in elucidating causal relations between parent characteristics and child outcomes. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 44(8), 1130-1144. doi:10.1111/1469-7610.00196
- Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, 32(1), 1-22.
- Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics*, 23(R1), R89-R98.
- Duffy, D. L., & Martin, N. G. (1994). Inferring the direction of causation in cross-sectional twin data - theoretical and empirical considerations *Genetic Epidemiology*, 11(6), 483-502. doi:10.1002/gepi.1370110606
- Eaves, L. J., Pourcain, B. S., Smith, G. D., York, T. P., & Evans, D. M. (2014). Resolving the effects of maternal and offspring genotype on dyadic outcomes in genome wide complex trait analysis ("M-GCTA"). *Behavior Genetics*, 44(5), 445-455.
- Fletcher, J. M. (2011). The promise and pitfalls of combining genetic and economic research. *Health economics*, 20(8), 889-892.
- Fletcher, J. M., & Lehrer, S. F. (2011). Genetic lotteries within families. *Journal of health economics*, 30(4), 647-659.
- Gage, S. H., Smith, G. D., Ware, J. J., Flint, J., & Munafò, M. R. (2016). G= E: What GWAS can tell us about the environment. *PLoS genetics*, 12(2).

- Gillespie, N. A., Gehrman, P., Byrne, E. M., Kendler, K. S., Heath, A. C., & Martin, N. G. (2012). Modeling the direction of causation between cross-sectional measures of disrupted sleep, anxiety and depression in a sample of male and female Australian twins. *Journal of sleep research, 21*(6), 675-683.
- Gillespie, N. A., & Martin, N. G. (2005). Direction of causation models. *Encyclopedia of statistics in behavioral science*.
- Gjerde, L. C., Eilertsen, E. M., Hannigan, L. J., Eley, T., Røysamb, E., Reichborn-Kjennerud, T., . . . Ystrom, E. (2019). Associations between maternal depressive symptoms and risk for offspring early-life psychopathology: the role of genetic and non-genetic mechanisms. *Psychological Medicine, 1*-9.
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102.
- Hannigan, L. J., Eilertsen, E. M., Gjerde, L. C., Reichborn-Kjennerud, T., Eley, T. C., Rijdsdijk, F. V., . . . McAdams, T. A. (2018). Maternal prenatal depressive symptoms and risk for early-life psychopathology in offspring: genetic analyses in the Norwegian Mother and Child Birth Cohort Study. *The Lancet Psychiatry, 5*(10), 808-815.
- Haworth, C. M., Nelson, S. K., Layous, K., Carter, K., Bao, K. J., Lyubomirsky, S., & Plomin, R. (2016). Stability and change in genetic and environmental influences on well-being in response to an intervention. *Plos one, 11*(5), e0155538.
- Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C. E., . . . Davis, O. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*(11), 1112.
- Heath, A. C., Kendler, K. S., Eaves, L. J., & Markell, D. (1985). The resolution of cultural and biological inheritance: Informativeness of different relationships. *Behavior Genetics, 15*(5), 439-465.
- 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. *Behavior Genetics, 23*(1), 29-50. doi:10.1007/bf01067552
- Hjelmborg, J., Korhonen, T., Holst, K., Skytthe, A., Pukkala, E., Kutschke, J., . . . Czene, K. (2017). Lung cancer, genetic predisposition and smoking: the Nordic Twin Study of Cancer. *Thorax, 72*(11), 1021-1027.
- Hur, Y.-M. (2008). Sex differences in genetic and environmental contributions to depression symptoms in South Korean adolescent and young adult twins. *Twin Research and Human Genetics, 11*(3), 306-313.
- Hur, Y.-M., Bogl, L. H., Ordonana, J. R., Taylor, J., Hart, S. A., Tuvblad, C., . . . Willemsen, G. (2020). Twin Family Registries Worldwide: An Important Resource for Scientific Research. *Twin Research and Human Genetics, 1*-11.
- Hwang, L.-D., Davies, N. M., Warrington, N. M., & Evans, D. M. (2020). Integrating Family-Based and Mendelian Randomization Designs. *Cold Spring Harbor Perspectives in Medicine*. doi:10.1101/cshperspect.a039503
- Jaffee, S. R., & Price, T. S. (2007). Gene-environment correlations: a review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry, 12*(5), 432-442. doi:10.1038/sj.mp.4001950
- Keller, M. C., Medland, S. E., Duncan, L. E., Hatemi, P. K., Neale, M. C., Maes, H. H. M., & Eaves, L. J. (2009). Modeling Extended Twin Family Data I: Description of the Cascade Model. *Twin Research and Human Genetics, 12*(1), 8-18. Retrieved from <Go to ISI>://WOS:000263635000002

- Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychological Medicine*, *37*(5), 615-626.  
doi:10.1017/s0033291706009524
- Kendler, K. S., Jacobson, K., Myers, J., & Eaves, L. (2008). A genetically informative developmental study of the relationship between conduct disorder and peer deviance in males. *Psychological Medicine*, *38*(7), 1001-1011.
- Kendler, K. S., Jacobson, K. C., Gardner, C. O., Gillespie, N., Aggen, S. A., & Prescott, C. A. (2007). Creating a social world: A developmental twin study of peer-group deviance. *Archives of General Psychiatry*, *64*(8), 958-965.
- Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics*, *23*(1), 21-27.
- Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsdottir, B. J., Young, A. I., Thorgeirsson, T. E., . . . Masson, G. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428.
- Kuja-Halkola, R., D'Onofrio, B. M., Larsson, H., & Lichtenstein, P. (2014). Maternal smoking during pregnancy and adverse outcomes in offspring: genetic and environmental sources of covariance. *Behavior Genetics*, *44*(5), 456-467.
- Lim, K. X., Liu, C. Y., Schoeler, T., Cecil, C. A., Barker, E. D., Viding, E., . . . Pingault, J. B. (2018). The role of birth weight on the causal pathway to child and adolescent ADHD symptomatology: a population-based twin differences longitudinal design. *Journal of Child Psychology and Psychiatry*, *59*(10), 1036-1043.
- Luo, Y. L., Haworth, C. M., & Plomin, R. (2010). A novel approach to genetic and environmental analysis of cross-lagged associations over time: The cross-lagged relationship between self-perceived abilities and school achievement is mediated by genes as well as the environment. *Twin Research and Human Genetics*, *13*(5), 426-436.
- Lynch, K. E. (2020). The meaning of "cause" in genetics. *Cold Spring Harbor Perspectives in Medicine*.
- Maes, H. H., Neale, M. C., Medland, S. E., Keller, M. C., Martin, N. G., Heath, A. C., & Eaves, L. J. (2009). Flexible Mx Specification of Various Extended Twin Kinship Designs. *Twin Research and Human Genetics*, *12*(1), 26-34. Retrieved from <Go to ISI>://WOS:000263635000004
- Malanchini, M., Wang, Z., Voronin, I., Schenker, V. J., Plomin, R., Petrill, S. A., & Kovas, Y. (2017). Reading self-perceived ability, enjoyment and achievement: A genetically informative study of their reciprocal links over time. *Developmental Psychology*, *53*(4), 698.
- Martin, N., Carr, A., Oakeshott, J., & Clark, P. (1982). Co-twin control studies: vitamin C and the common cold. *Prog. Clin. Biol. Res*, *103*, 365-373.
- McAdams, T. A., Gregory, A. M., & Eley, T. C. (2013). Genes of Experience: Explaining the Heritability of Putative Environmental Variables Through Their Association with Behavioural and Emotional Traits. *Behavior Genetics*, *43*(4), 314-328.  
doi:10.1007/s10519-013-9591-0
- McAdams, T. A., Hannigan, L. J., Eilertsen, E. M., Gjerde, L. C., Ystrom, E., & Rijdsdijk, F. V. (2018). Revisiting the children-of-twins design: improving existing models for the exploration of intergenerational associations. *Behavior Genetics*, *48*(5), 397-412.

- McAdams, T. A., Neiderhiser, J. M., Rijdsdijk, F. V., Narusyte, J., Lichtenstein, P., & Eley, T. C. (2014). Accounting for genetic and environmental confounds in associations between parent and child characteristics: A systematic review of children-of-twins studies. doi:10.1037/a0036416
- Millwood, I. Y., Walters, R. G., Mei, X. W., Guo, Y., Yang, L., Bian, Z., . . . Hu, R. (2019). Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and women in China. *The Lancet*, *393*(10183), 1831-1842.
- Narusyte, J., Neiderhiser, J. M., D'Onofrio, B. M., Reiss, D., Spotts, E. L., Ganiban, J., & Lichtenstein, P. (2008). Testing Different Types of Genotype-Environment Correlation: An Extended Children-of-Twins Model. *Developmental Psychology*, *44*(6), 1591-1603. doi:10.1037/a0013911
- Neiderhiser, J. M., Reiss, D., & Hetherington, E. M. (2007). The Nonshared Environment in Adolescent Development (NEAD) Project: A longitudinal family study of twins and siblings from adolescence to young adulthood. *Twin Research and Human Genetics*, *10*(1), 74-83.
- Neiderhiser, J. M., Reiss, D., Lichtenstein, P., Spotts, E. L., & Ganiban, J. (2007). Father-adolescent relationships and the role of genotype-environment correlation. *Journal of Family Psychology*, *21*(4), 560-571. doi:10.1037/0893-3200.21.4.560
- Neiderhiser, J. M., Reiss, D., Pedersen, N. L., Lichtenstein, P., Spotts, E. L., & Hansson, K. (2004). Genetic and environmental influences on mothering of adolescents: A comparison of two samples. *Developmental Psychology*, *40*(3), 335-351. doi:10.1037/0012-1649.40.3.335
- Ozaki, K., & Ando, J. (2009). Direction of causation between shared and non-shared environmental factors. *Behavior Genetics*, *39*(3), 321-336.
- Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijdsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, *19*(9), 566.
- Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science*, *11*(1), 3-23.
- Polderman, T. J., 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. *Nature genetics*, *47*(7), 702.
- Rice, F., van den Bree, M. B. M., & Thapar, A. (2004). A population-based study of anxiety as a precursor for depression in childhood and adolescence. *BMC Psychiatry*, *4*(1), 43. doi:10.1186/1471-244X-4-43
- Richmond, R. C., & Davey Smith, G. (2020). Mendelian Randomization: Concepts and Scope. *Cold Spring Harbor Perspectives in Medicine*.
- Selzam, S., McAdams, T. A., Coleman, J. R., Carnell, S., O'Reilly, P. F., Plomin, R., & Llewellyn, C. H. (2018). Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. *PLoS genetics*, *14*(11), e1007757.
- Selzam, S., Ritchie, S. J., Pingault, J.-B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing within-and between-family polygenic score prediction. *BioRxiv*, 605006.
- Silberg, J. L., Copeland, W., Linker, J., Moore, A. A., Roberson-Nay, R., & York, T. P. (2016). Psychiatric outcomes of bullying victimization: a study of discordant monozygotic twins. *Psychological Medicine*, *46*(9), 1875-1883.

- Silberg, J. L., & Eaves, L. J. (2004). Analysing the contributions of genes and parent-child interaction to childhood behavioural and emotional problems: a model for the children of twins. *Psychological Medicine*, *34*(2), 347-356. doi:10.1017/s0033291703008948
- Silberg, J. L., Maes, H., & Eaves, L. J. (2010). Genetic and environmental influences on the transmission of parental depression to children's depression and conduct disturbance: an extended Children of Twins study. *Journal of Child Psychology and Psychiatry*, *51*(6), 734-744. doi:10.1111/j.1469-7610.2010.02205.x
- Singham, T., Viding, E., Schoeler, T., Arseneault, L., Ronald, A., Cecil, C. M., . . . Pingault, J.-B. (2017). Concurrent and longitudinal contribution of exposure to bullying in childhood to mental health: the role of vulnerability and resilience. *Jama Psychiatry*, *74*(11), 1112-1119.
- Smith, G. D. (2011). Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *International Journal of Epidemiology*, *40*(3), 537-562. doi:10.1093/ije/dyr117
- Smith, G. D., & Ebrahim, S. (2005). What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *Bmj*, *330*(7499), 1076-1079.
- Tick, B., Colvert, E., McEwen, F., Stewart, C., Woodhouse, E., Gillan, N., . . . Simonoff, E. (2016). Autism Spectrum Disorders and other mental health problems: Exploring etiological overlaps and phenotypic causal associations. *Journal of the American academy of child & adolescent psychiatry*, *55*(2), 106-113. e104.
- Tikhodeyev, O. N., & Shcherbakova, O. V. (2019). The problem of non-shared environment in behavioral genetics. *Behavior Genetics*, *49*(3), 259-269.
- Torvik, F. A., Eilertsen, E. M., McAdams, T. A., Gustavson, K., Zachrisson, H. D., Brandlistuen, R., . . . Ask, H. (2020). Mechanisms linking parental educational attainment with child ADHD, depression, and academic problems: a study of extended families in The Norwegian Mother, Father and Child Cohort Study. *Journal of Child Psychology and Psychiatry*.
- Torvik, F. A., Rosenström, T. H., Gustavson, K., Ystrom, E., Kendler, K. S., Bramness, J. G., . . . Reichborn-Kjennerud, T. (2019). Explaining the association between anxiety disorders and alcohol use disorder: A twin study. *Depression and Anxiety*, *36*(6), 522-532.
- Torvik, F. A., Rosenström, T. H., Ystrom, E., Tambs, K., Røysamb, E., Czajkowski, N., . . . Reichborn-Kjennerud, T. (2017). Stability and change in etiological factors for alcohol use disorder and major depression. *Journal of Abnormal Psychology*, *126*(6), 812.
- Truett, K. R., Eaves, L. J., Walters, E. E., Heath, A. C., Hewitt, J. K., Meyer, J. M., . . . Kendler, K. S. (1994). A model system for analysis of family resemblance in extended kinships of twins *Behavior Genetics*, *24*(1), 35-49. doi:10.1007/bf01067927
- Turkheimer, E., & Harden, K. P. (2014). Behavior genetic research methods. *Handbook of research methods in social and personality psychology*, 159-187.
- Tuvblad, C., Grann, M., & Lichtenstein, P. (2006). Heritability for adolescent antisocial behavior differs with socioeconomic status: gene–environment interaction. *Journal of Child Psychology and Psychiatry*, *47*(7), 734-743.
- Verhulst, B. (2017). A power calculator for the classical twin design. *Behavior Genetics*, *47*(2), 255-261.
- Visscher, P. M. (2004). Power of the classical twin design revisited. *Twin Research and Human Genetics*, *7*(5), 505-512.

- Xu, C., Sun, J., Ji, F., Tian, X., Duan, H., Zhai, Y., . . . Zhao, Z. (2015). The genetic basis for cognitive ability, memory, and depression symptomatology in middle-aged and elderly Chinese twins. *Twin Research and Human Genetics*, *18*(1), 79-85.
- Young, A. I., Benonisdottir, S., Przeworski, M., & Kong, A. (2019). Deconstructing the sources of genotype-phenotype associations in humans. *Science*, *365*(6460), 1396-1400.
- Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., . . . Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature genetics*, *50*(9), 1304.
- Ystrom, E., & Eilertsen, E. M. (2019). Deconstructing the history of genotype-phenotype associations in humans.