Gage, S. H., Munafo, M. R., & Davey Smith, G. (2016). Causal inference in Developmental Origins of Health and Disease (DOHaD) research. *Annual Review of Psychology*, *67*, 567-585. https://doi.org/10.1146/annurev-psych-122414-033352

# Causal Inference in Developmental Origins of Health and Disease (DOHaD) Research

Suzanne H Gage [1,2], Marcus R Munafò [1,2], George Davey Smith [1]

1. MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, United Kingdom
2. UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

Corresponding author email: George Davey Smith, MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, United Kingdom. Email: Julia.Mackay@bristol.ac.uk

Abstract

Studies of the Developmental Origins of Health and Disease (DOHaD) often rely on prospective observational data, from which associations between developmental exposures and outcomes in later life can be identified. Typically, conventional statistical methods are used in an attempt to mitigate problems inherent in observational data, such as confounding and reverse causality, but these have serious limitations. In this review we discuss a variety of methods that are increasingly being used in observational epidemiological studies to help strengthen causal inference. These include negative controls, cross-contextual designs, instrumental variables (including Mendelian randomization), family-based studies and natural experiments. Applications within the DOHaD framework, and in relation to behavioural, psychiatric and psychological domains, are considered, and the considerable potential for expanding the use of these methods is outlined.

Keywords: DOHaD; Causal Inference; Instrumental Variable; Negative Control; Cross-Contextual Comparison; Twin Study.

## Contents

Causal Inference in Developmental Origins of Health and Disease (DOHaD) Research

## Introduction

The Developmental Origins of Health and Disease (DOHaD) hypothesis proposes that the environment an individual experiences in utero and during early development can affect their health and susceptibility to disease over the rest of their life. A long lineage can be traced for what is now referred to as DOHaD (Kuh and Davey Smith, 2004), but contemporary interest increased following work from David Barker and colleagues in the mid-1980s that suggested that early life nutrition is associated with cardiovascular disease risk in later life (Barker, 1995, Fall et al., 1995, Barker and Osmond, 1986). The concept that early life experiences have long term effects is not new to psychology. From imprinting in Lorenz's geese (Lorenz, 1935) to the long-term effects of trauma on little Albert (Watson and Rayner, 1920), theories conceptually similar to the DOHaD hypothesis have played a key role in psychological research for many decades. Understanding these relationships is important, since elucidating the mechanisms by which early life experiences can affect adult physical and mental well-being will identify potentially important targets for intervention to prevent adverse outcomes from occurring, many years before they are likely to do so.

There is a substantial body of evidence in support of the DOHaD hypothesis. Barker and colleagues' original series of papers presented evidence of associations between low birth weight and a number of offspring health outcomes, including risk of coronary heart disease and stroke (Barker, 1997), hypertension (Barker and Martyn, 1997), and type-2 diabetes (Hales and Barker, 1992). These studies were originally considered under the umbrella term of 'the fetal origins of adult disease' (FOAD), as the major focus was on the role of the intrauterine environment on later offspring

outcomes, but the concept was later extended to include other aspects of developmental plasticity, including the early post-natal period and possible pre-conceptual and intergenerational influences. Since the initial papers were published, low birth weight has also been shown to be associated with offspring obesity (Eriksson et al., 2015), depression (Van Lieshout and Boylan, 2010) and intelligence (Eryigit Madzwamuse et al., 2014). However, studies to date have for the most part been conducted using observational data; therefore, while they provide suggestive evidence that these developmental influences affect later outcomes, they are limited in terms of providing strong enough evidence that a causal interpretation can be drawn.

The purpose of this review is to describe the limitations of traditional methods for assessing associations in observational studies and inferring causality, and to provide an introduction to alternative approaches to fashioning and analyzing data sources that can help in this regard. Although DOHaD is the main lens through which these questions will be discussed, we extend it to consider "development" and health more generally, as there are other time periods that are likely to be critical for later physical and mental health, such as adolescence, for which the same issues apply.

## Problems with Observational Studies

When considering the impact of an exposure on an outcome, the strongest evidence of a causal association comes from experimental designs, in particular randomised controlled trials (RCTs), where a group of individuals are randomly assigned to either an exposure condition or a control condition, and followed up to ascertain differential incidence of the outcome between the two groups. However, the use of an experimental design is not possible when it is unethical or impractical to either give or withhold a particular exposure. For example, when the exposure in question might only have an effect after many years, such RCTs would be

prohibitively expensive and impractical to run. Therefore, in order to attempt to ascertain causation in these circumstances, observational data must be interrogated. Without the ability to randomize people, associations seen between an exposure and outcome in observational data could be due to a number of other possibilities, aside from a causal association between the two. These include confounding, reverse causation and various biases that could distort the underlying association (Davey Smith and Ebrahim, 2002).

*Confounding.* If there are other differences between those who experience the exposure and those who do not, any association seen could be due to these confounding factors. Adjustment must be made in analyses for all potential confounders. However, statistical adjustment will usually be incomplete: not only must all the confounders be measured, but confounders must suffer from no measurement error for such adjustments to successfully account for confounding. Unmeasured confounding factors and measurement error (due to either technical issues or to temporal variation in a factor assessed only once) in assessed confounding factors leads to *residual confounding* even when confounders have apparently been statistically "controlled" for in the analysis (Fewell et al., 2007). For this reason, residual confounding can never be completely ruled out in observational studies.

*Reverse causation*. When assessing observational data, it is challenging to ascertain the direction of causation, even when there is a temporal gap between exposure and outcome. Pre-existing symptoms of the outcome which influence the exposure could generate the observed associations. For example, observational evidence has shown that alcohol consumption is associated with mortality in a J-shaped curve, with those who drink nothing at all showing worse outcomes than those who drink a small amount. It has been suggested that this association might be seen because some non-drinkers stop drinking due to ill health; the drinking

behaviour is a *consequence* of the increased risk, rather than the other way round (Liang and Chikritzhs, 2013).

*Selection Bias.* Estimates seen in observational studies can be affected by selection bias because of how participants are recruited into a study, or how data from participants are collected. For example, certain types of people might be more likely to be lost to follow-up in longitudinal studies. If loss to follow-up is related to two or more variables then the available sample is, in effect, stratified by whether follow-up was successful or not, which generates associations between these variables in the available dataset even when associations do not exist in the underlying population, and could change the strength and even direction of associations that do exist. This is a form of collider bias (Cole et al., 2010) - a family of biases that can distort observational estimates of exposure effects - that have perhaps been under-appreciated in the literature until recently.

Misclassification can occur when participants are incorrectly assigned to an exposure or outcome category due to imprecise data collection methods, and if this is differential (e.g., degree of misclassification of outcome relates to the exposure), it can distort exposure-outcome associations. Self-report measures might not adequately capture variables when participants might want to hide their use of a substance (e.g., smoking during pregnancy). Such information biases may particularly influence case-control studies when retrospective reporting of exposures occurs after the outcome condition has developed (Rothman et al., 2008).

The usual approach to attempting to mitigate the potential biases above is to use statistical methods aimed at removing or minimising them. However, statistical analyses necessarily require assumptions to be made about the data, and these may be inappropriate or inadequate with respect to the structure of the data. Moreover, statistical adjustments can lead to over-confidence in the robustness of findings – the commonly used term that factors have been statistically “controlled for” gives a sense of this – leading to the literature containing many associations that are over-

interpreted in terms of causal evidence (Davey Smith and Ebrahim, 2001). As with residual confounding, attempted statistical adjustment to account for potential biases has serious limitations.

In this review, we discuss alternative approaches to conventional statistical adjustment methods which can improve causal inference. Ways of assembling and analysing data can strengthen such inference, and triangulating evidence from multiple independent sources can provide more reliable evidence for causation than a single approach. While experimental designs are typically not ethical or practical, identifying study designs that either reveal bias, confounding or reverse causality, or are better protected from these than conventional approaches, and applying these to questions related to the DOHaD hypothesis, will provide a far stronger foundation for this literature (Richmond et al., 2014).

## Methods for Causal Inference

Below we describe methods from epidemiological studies that attempt to address problems of confounding, reverse causation and bias at the *design* stage of a study, rather than relying on statistical methods after data collection. We argue that such methods allow for stronger causal inference, and have the potential to provide much stronger evidence to elucidate the mechanisms that might underlie the associations between developmental experiences and adult physical and mental health. Table 1 summarises these methods.

Insert Table 1 about here.

### Negative Controls

When assessing an observational association, one cannot be certain whether the association being seen is due to residual confounding. One method to examine this possibility is to compare the association of interest with that of another related

association, but for which there is no biologically plausible mechanism for causation. This is known as a “negative control” design, and was developed in the economics and econometrics literature (DiNardo and Pischke, 1997, Oosterbeek, 1997). The negative control analysis will have either the same exposure or the same outcome as in the main analysis of interest, but will replace either the exposure or the outcome with a “negative control” for this, in order to uncover potential unobserved or unaccounted-for confounding or bias. A suitable negative control should be subject to the same confounding structure as the association of interest. Associations between the exposure and outcome of interest are then compared to those between the negative control exposure and the outcome of interest, or the exposure of interest and negative control outcome. If there is a causal association between the exposure and outcome of interest, it would be expected that there would be no association, or a considerably smaller association, in the negative control analysis. If, however, the association is due to confounding, then a similar association is likely to be seen in both the analysis of interest, and the negative control analysis, where there is no biologically plausible mechanism for causation (Davey Smith, 2012, Lipsitch et al., 2010). Figure 1 shows an example of a negative control exposure, and a negative control outcome design.

Insert Figure 1 about here.

The rationale behind negative control designs is that the inspection of analyses that utilize negative control exposures or outcomes – which are likely to share similar confounding with the exposure or outcome of interest - can help strengthen causal inference. For example, consider maternal smoking during pregnancy, which researchers have hypothesised may cause offspring depression through a direct intrauterine effect. A plausible negative control exposure in this situation is paternal smoking during pregnancy, where no substantial intrauterine

biological effect will occur, but confounding factors are likely to be similar. Researchers can also investigate the relationships between an exposure and a negative control outcome. The negative control outcome should be influenced by similar confounding and other biases as would be seen for the outcome of interest, but be unlikely to be caused by the exposure. For example, as smoking has similar associations with both suicide and homicide mortality, this casts doubt on smoking *causing* suicide; whilst there are apparently plausible causal biological mechanisms that can be advanced to explain the smoking-suicide association, the same is not the case for the smoking-homicide association (Davey Smith et al., 1992).

*Negative control exposures.* Negative control exposures have been used in studies trying to assess the potential causal effects of periconceptual folate or folic acid supplementation. Given the established causal association between inadequate periconceptual folate status and neural tube defects (Pitkin, 2007), randomized trials deliberately withholding advice to take periconceptual folate supplements from the control group would be unethical, so trials aimed at evaluating the effect on other outcomes are unlikely to be undertaken. The observational associations seen between folate supplement use and other outcomes such as increased rates of autism spectrum disorders or slower language development could be due to residual confounding from socio-economic position or health-adverse maternal behaviours in general (Davey Smith, 2008). In order to strengthen causal inference with regard to maternal periconceptual folate supplementation and autism, one study examined the association between fish oil supplements and autism in the same sample (Suren et al., 2013), since the use of fish oil supplements and use of folic acid supplements were similarly socially patterned with respect to potential confounders such as parental characteristics. There was a robust inverse association between use of folic acid supplements and subsequent risk of autism spectrum disorder. However, there was little evidence of an association between use of fish oil supplements and autism spectrum disorder. The difference between these two suggests that it is not simply

residual confounding from maternal health-related behaviours or social circumstances more generally that is leading to the observed association between maternal folate and autism spectrum disorders.

Similarly, another study assessed the association between folate supplementation and language delay (Roth et al., 2011). This group used a four-category exposure measure of ‘no supplement’, ‘supplements other than folic acid’, ‘only folic acid’, and ‘folic acid plus other supplements’. The authors found that there was little evidence of an association between supplements other than folic acid and later language delay, compared to the baseline of ‘no supplement use’, despite the different supplements showing similar associations with confounding factors. However, an inverse association was seen for both of the groups in which the different supplements included folic acid. This provided further evidence that the association seen with folate supplementation could be a protective one, and not due to residual confounding - although of course causation is still not in any sense “proven”.

As already introduced above, a now widely used negative control exposure for studies investigating effects thought to occur in utero is to examine the same association for exposures in fathers, rather than mothers, since a direct intrauterine effect will not occur in the former case. Brion and colleagues found that maternal macronutrient and energy intake during pregnancy predicted later offspring dietary intake, while paternal nutrition during partners’ pregnancy did not. They concluded that this provides some evidence that maternal intake during pregnancy could programme later offspring appetite (Brion et al., 2010a). Conversely, associations between maternal or paternal smoking and later offspring blood pressure were similar, suggesting that the association seen is unlikely to be due to an intrauterine effect and could indicate that residual confounding is impacting on the associations seen (Brion et al., 2007).

*Negative control outcomes.* Negative control outcomes use broadly the same principles as negative control exposures. An outcome variable is selected which is unlikely to be caused is by the exposure of interest.

An example of negative control outcomes is taken from studies of hormone replacement therapy (HRT). Many early studies found evidence that HRT was associated with lower mortality from cardiovascular disease. In the late 1980's Petitti and colleagues similarly found evidence that HRT was associated with lower mortality from cardiovascular disease, a result they described as 'suggestive'. However, they conducted a further analysis to also assess rates of mortality from accidents, suicide and homicide in women using HRT compared to those not using it, where there is no plausible biological mechanism (Petitti et al., 1987, Petitti et al., 1986). They found evidence that HRT was associated with lower rates of these forms of mortality as well, and suggested that this finding indicated that at least some of the differences in outcomes seen between HRT users and non-users were likely to be due to lifestyle, socioeconomic, behavioural and related differences. As was later borne out by RCTs, the observational evidence suggesting that HRT substantially reduced the risk of cardiovascular mortality was indeed spurious (Lawlor et al., 2004).

*Limitations.* The use of negative controls can provide useful evidence of residual confounding, if similar associations are seen for the negative control exposure or in the negative control outcome. However, if associations are not similar between the association of interest and that seen in the negative control, this is not of course definitive proof of causation, as the association of interest could still be confounded by other factors that are not shared with the negative control, or be subject to bias. The technique is also inappropriate if there might be a plausible biological mechanism impacting on the negative control. For example, paternal smoking during pregnancy could conceivably affect the developing foetus via the effects of environmental tobacco smoke exposure (Taylor et al., 2014) although the

association with outcomes of interest would still be expected to be attenuated relative to associations observed with maternal smoking during pregnancy.

## Cross-Contextual Comparisons

Cross-contextual comparisons operate on the opposite principle to negative control methods. These designs look for similar associations in very different populations (typically across different countries). If the same association between exposure and outcome is seen in populations where the underlying confounding structures are very different, this provides stronger evidence of causality. It would mean that if the association seen was due to residual confounding it would have to be from different sources in the different populations, which is not likely. Figure 2 illustrates this.

Insert Figure 2 about here.

One way this design has been utilized is by comparing birth cohorts in different countries. Brion and colleagues (Brion et al., 2011) used the Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort based in the United Kingdom, and the Pelotas cohort based in Brazil to assess the causal effects of breastfeeding on various outcomes. While breastfeeding in the United Kingdom is associated with higher socio-economic position, healthier diet, and lower levels of maternal smoking, there is not the same social patterning of breastfeeding behaviour in Brazil. Whereas in the ALSPAC cohort the authors found that breastfeeding was associated with lower offspring body mass index and blood pressure, these inverse associations were not seen in the Pelotas cohort. The authors used this divergence in cross-contextual findings to provide suggestive evidence that many of the associations seen in observational studies in Western countries between breastfeeding and various outcomes are likely to be due to residual confounding.

They found some evidence that the association between breastfeeding and IQ might be causal, and indeed this has been supported by results from a RCT conducted in Belarus (Kramer et al., 2007, Patel et al., 2014), and from further recent evidence from a different wave of the Pelotas cohort in Brazil (Victora et al., 2015).

A second example of cross-contextual comparison would be to compare cohorts where the same exposure has changed over time. For example, breastfeeding was not strongly socially patterned in the United Kingdom in the 1920s. Martin and colleagues (Martin et al., 2007) compared bottle-fed and breastfed infants in the Boyd-Orr Survey of Diet and Health who were born in the 1920s and 1930s, and surveyed between 1937 and 1939, and again between 1997 and 1998. They found that breastfeeding in the 1920s was associated with upward social mobility (i.e., odds of moving from a lower to a higher occupational social class from childhood to adulthood). Critically, breastfeeding was not associated with indicators of socioeconomic position such as household income. Therefore consistency in associations of breastfeeding with offspring health outcomes over time, as confounding structures have changed, increases confidence that these associations are causal in nature.

On the other hand, if associations change over time, this suggests that the association may *not* be causal. An example of this is cannabis research. There is some evidence that cannabis use during adolescence could be damaging to later health due to changes that occur in the endocannabinoid system during this period of development (Rubino and Parolaro, 2008, Trezza et al., 2008). Levels of tetrahydrocannabinol (THC) and other cannabinoids in “street” cannabis have changed substantially since investigations into the association between cannabis and psychosis were first conducted (Mehmedic et al., 2010). If the nature of the association between cannabis and psychosis is a biological effect of THC, it might be expected that associations between cannabis and psychosis seen would be stronger in more recent cohort studies than in earlier ones. However, the association between

cannabis use and psychosis was first reported in a cohort measured in the 1960s (Zammit et al., 2002), which was before levels of THC are thought to have increased (King et al., 2004). The size of the point estimates seen in individual studies has not increased since this first study in a systematic way. A recent case-control study suggested that “skunk” cannabis (high in THC but with little cannabidiol) was associated with hospitalization for first-episode psychosis, but “hash” cannabis (with equivalent levels of THC and CBD) was not (Di Forti et al., 2015). While this could provide evidence in support of an effect of cannabis strength on risk of psychosis, an alternative explanation might be that people at a higher risk of psychosis for other reasons are self-selecting to an extreme end of a distribution, which in the past might have been using cannabis more frequently, but as stronger strains have become available has changed to higher potency of cannabis (Gage et al., 2015).

*Limitations.* When conducting such cross-contextual studies, it is important to ascertain whether there are relevant differences in confounding structure between the two populations being compared. If there are still similar confounders in both contexts that could be driving the association seen, then the comparison is inappropriate. Also, the exposure and outcome variables being compared need to be harmonized across the cohorts in order to be directly comparable. When an exposure has changed over time, a number of other variables might have also changed which could confound the association.

## Instrumental Variable Analyses

Another method to help strengthen causal inference in observational data was conceived in the econometrics literature. Instrumental variable analyses use a proxy variable (known as an “instrumental variable” or instrument) in place of the exposure of interest. If an appropriate instrument can be identified, it should in principle allow for causal interpretation from observational data. However, the proposed instrument must satisfy three assumptions be a valid instrumental variable.

First, it should be robustly associated with the exposure of interest. Second, it should not associate with potential confounding factors, either known or unknown, that can bias naïve observational associations. Third, it should not directly affect the outcome of interest (Angrist and Pischke, 2009). A diagram of these requirements is shown in Figure 3a.

One study used a short-lived policy change, which had unintended consequences, in an instrumental variable design. In Sweden, a law was introduced that made strong beer easily purchasable by those under 21, when previously it had not been. This policy change was used as a proxy for *in utero* alcohol exposure. Critically, the participants in the study were conceived prior to the policy being introduced. So the pregnancies were not due to an increase in unplanned pregnancies resulting from risky sexual behaviours following increased alcohol consumption. The author of the study found that children born to mothers under 21 who were pregnant for the longest period during the policy change (5 to 8.5 months) had lower earnings and wages, were more likely to be unemployed, and had higher welfare dependency rates, compared to cohorts from other parts of Sweden, or those occurring just before or just after the policy change (Nilsson, 2014).

*Limitations.* The principal limitation of instrumental variable methods is that it can be challenging to identify valid instruments that are genuinely not associated with potential confounders, and not subject to reverse causality. Critically, it is not possible to definitively test the validity of putative instruments, since unmeasured confounders may be operating.

Insert Figure 3 about here.

## Mendelian Randomization

Mendelian randomization (MR) is a type of instrumental variable analysis which uses using genetic variants as unconfounded proxies (i.e., instruments) for the

exposure of interest. Due to the random nature of inheritance of genetic information, it can be reasonably assumed that we inherit each variant (for the most part) independently from other genetic variants and from environmental factors, meaning such variants are unlikely to be associated with potential confounding factors. Also, because our genomes are determined at conception, associations between genetic variants and outcomes cannot be due to reverse causation. Therefore, if a genetic variant is robustly associated with an exposure of interest, it could potentially be used in a Mendelian randomization experiment (Davey Smith, 2010, Davey Smith and Ebrahim, 2003). This is illustrated in Figure 3b. With regards to developmental outcomes, there are single nucleotide polymorphisms (SNPs) or genetic risk scores already identified via genome wide association studies that can be used as proxies for exposures such as smoking or drinking during pregnancy, or for maternal body mass index.

A genetic variant has been identified that robustly correlates to smoking heaviness in daily smokers (Ware et al., 2012). Located in the *CHRNA5-A3-B4* gene cluster, on chromosome 15, rs1051730 and rs16969968 are in perfect linkage disequilibrium and can be treated as interchangeable. Each additional copy of the minor (T) allele is associated with one extra cigarette smoked per day in smokers (Thorgeirsson et al., 2008), accounting for ~1% of the variation in cigarette consumption in daily smokers (Ware et al., 2011) and ~4% of cotinine levels, the primary metabolite of nicotine and a more precise biomarker of exposure (Munafo et al., 2012, Keskitalo et al., 2009). Crucially for investigating developmental outcomes, the variant has also been shown to associate with lack of ability to give up smoking during pregnancy (Freathy et al., 2009). This variant has been used in a number of Mendelian randomisation designs, including as a proxy for foetal exposure to cigarette smoke. Tyrrell and colleagues have shown that variation at this locus not only predicts an increased likelihood to continue smoking during pregnancy, but also a larger number of cigarettes per day in pregnant women who continue to smoke

(Tyrrell et al., 2012). The authors performed a meta-analysis of 14 studies, comprising 26,241 women. Of those who smoked beyond the first trimester during pregnancy, each additional copy of the rs1051730 T allele, associated with increased smoking, was associated with a 20g reduction in offspring birth weight. Conversely, they found little evidence of differences in birthweight by genotype in non-smokers. Given the genotype's lack of association with factors that usually confound observational associations such as age, socio-economic position, and occupation, and the lack of possibility of reverse causation in this type of design, this study provides much stronger evidence of causation than is possible from observational designs.

Genetic variants have also been identified that predict alcohol use (Enomoto et al., 1991). Although one variant is only prevalent in East Asian populations, there are also variants present in Western populations that can be used as a proxy for exposure to alcohol during pregnancy. Many observational studies have suggested a "J-shaped" association between alcohol use and many outcomes, where those who drink a small amount have better outcomes than those who do not drink at all. However, drinking behaviour is highly socially patterned, so residual confounding could well still be affecting these findings. For example, Kelly and colleagues (Kelly et al., 2012) reported that low levels of maternal alcohol consumption in pregnancy (1 to 2 drinks per week or per occasion) were associated with reduced behavioural difficulties and hyperactivity in offspring at age 5 years. However, data from the same study indicated that similar associations were observed for tobacco use and maternal socioeconomic position. Never drinking mothers and those who did not drink during pregnancy were more likely to smoke and more likely to have never worked or been long-term unemployed than light drinkers (see Figure 4). Also, reverse causation is harder to rule out in this context as people may have stopped drinking due to ill health, which might not be captured. Zuccolo and colleagues (Zuccolo et al., 2013) found evidence that a genotype associated with lower alcohol consumption or

abstinence during early pregnancy was associated with offspring higher school academic achievement at age 11, suggesting that alcohol exposure in utero is causally associated with lower offspring educational outcomes. However, they found little evidence of an association with childhood IQ at age 8, although the statistical power was lower for this analysis.

Insert Figure 4 about here.

Genetic variants that predict adiposity have been used as a proxy for maternal body mass index in order to ascertain potential programming effect of prenatal maternal obesity on offspring outcomes. Lawlor and colleagues (Lawlor et al., 2008) found little evidence that maternal genotype predicted offspring fat mass by age 9-11 years, after adjustment for offspring genotype (which is important when there could be a direct effect of offspring genotype on the outcome of interest). This suggests that the association between maternal and offspring adiposity may not operate via the prenatal environment.

*Limitations.* There are a number of circumstances in which Mendelian randomization is an inappropriate study design. Most obviously, if there is no genetic variant yet identified which is robustly associated with the exposure of interest, this design is not possible to use. For example, although cannabis use is known to be heritable, as yet genome wide association studies have not identified any variants robustly associated with cannabis use phenotypes. Given that the size of associations seen between variants and exposures of interest are often of modest size, large sample sizes are required for adequate power to undertake such study designs, meaning consortia are often necessary. This can lead to heterogeneous measures of the outcome as studies are combined. The most fundamental limitation relates to when a genetic variant has a direct pleiotropic effect (whereby a gene influences more than one phenotype) on the outcome of interest as well as the

exposure, as this can lead to spurious associations. Some genetic variants are in linkage disequilibrium, meaning they are more likely to be inherited together, which can generate similar biases as seen where there is pleiotropy. Methods to evaluate and account for such reintroduced confounding by pleiotropy or linkage disequilibrium are discussed elsewhere (Bowden et al., 2015, Davey Smith and Hemani, 2014). If the sample contains two or more ancestrally different populations, associations between genetic variants and outcomes could be due to population stratification, rather than being due to a causal effect of the exposure on the outcome. This can be accounted for by study designs utilising ancestrally similar samples, and by using principal components analysis to correct for stratification and account for ancestry (Davey Smith and Hemani, 2014).

### Family Design Techniques

The use of genetically-related pairs or groups of individuals can mean that potential confounding from genetics and shared environments (factors that are the same as the pairs grow up together) are less plausible explanations of observed associations (D'Onofrio et al., 2014). Twin and sibling designs have been used for many years, and there are a number of different possible designs for such studies. The classic twin study design compares monozygotic and dizygotic twin pairs in order to separate genetic (as this is correlated at 1.0 in monozygotic twins and 0.5 in dizygotic twins), shared environment (which correlated in both types of twin at 1.0) and non-shared environment (not correlated in either type of twin) influences.

If pairs of monozygotic twins reared together are discordant on the exposure of interest, they can be used as ideally matched pairs in a case-control study. Any association seen will not be due to confounding from genetic factors, and will not be due to confounding from shared environmental factors. However, this design cannot rule out the impact of non-shared environmental confounders. It is also important to consider that the intrauterine environment of a twin is not the same as that of a

singleton pregnancy, and this could mean results from such studies are less easily generalizable. Other designs using genetically-related individuals include using sibling or cousin pairs, which remove the shared intrauterine environment, but do not account for all genetic variation.

Twin studies have been used to investigate aspects of the DOHaD hypothesis (D'Onofrio et al., 2014). For example, Class and colleagues (Class et al., 2014) used a sibling comparison design approach to disentangle genetic and environmental effects on associations between foetal growth and psychiatric and socio-economic problems. The found that, within sibling pairs, lower birth weight predicted autism spectrum disorder and attention deficit hyperactivity disorder. However, when they assessed associations with suicide attempt and substance use, these associations were fully attenuated in sibling comparison models where the sibling differed in their substance use, suggesting that evidence of such an association seen in more traditional cohort designs where non-related individuals are randomly sampled may have been due to residual confounding.

Another study used dizygotic female twins in a study to investigate whether exposure to testosterone in utero increases the risk of attention deficit hyperactivity disorder and autism spectrum disorder (Attermann et al., 2012). The sex of the participants' co-twin was used as a genetic proxy for exposure to testosterone, as a male co-twin would increase the female twin's exposure to prenatal testosterone. However, sex of the co-twin should not be confounded with other genetic or environmental factors. The authors found that having a male co-twin was associated with a reduction of risk of attention deficit hyperactivity disorder and autism spectrum disorder, opposite to what had been predicted. They concluded that this could be due to parental reporting bias, or unmeasured variables still confounding the association.

*Limitations.* Depending on the type of twin or family design employed, there are different limitations to these study designs. Most notably, finding monozygotic twins discordant on the exposure of interest is not trivial, which can make these

studies challenging to conduct, or result in them being underpowered. The potential lack of generalisability due to different intrauterine experience from twin versus singleton births can also limit the impact of some of these study designs, particularly when assessing exposures occurring during that time.

## Natural Experiments

Occasionally, situations will arise whereby unusual circumstances can provide insights that observational studies cannot. One such event was the Dutch Hunger Winter of 1944-1945. Towards the end of World War II, a Nazi blockade led to a severe food shortage in Netherlands, where civilians were subjected to rations equivalent to less than 500 calories per day. Pregnancies that occurred during this period represent a unique opportunity to experimentally investigate the impact of severe calorific restriction upon offspring outcomes.

Early studies using the cohort found an association between conception during the height of the famine and neural tube defects (spina bifida and anencephaly), compared to the background rate of such disorders in the Dutch population (Brown and Susser, 1997). When the cohort were older, an association with schizophrenia was also assessed, which found the cumulative risk of schizophrenia between 24 and 48 years old to be double that of unaffected comparison cohorts, and even compared to those exposed to the famine during other periods of gestation (Hoek et al., 1998). This finding was replicated in another natural experiment that was possible after a famine in China brought about by the Great Leap Forward period of social and economic upheaval. Although there was not caloric intake data available for this cohort, it was still possible to assess the impact of famine during conception on risk of later schizophrenia. The impacts of severe caloric restriction in this very culturally different cohort were largely similar to those seen in Netherlands (Song et al., 2009).

*Limitations.* Such extreme events as famine may have other consequences that could confound associations seen. For example, prenatal stress is likely to have been much higher during these periods than surrounding times. However, in the Dutch study it was possible to compare with cohorts in other areas of Netherlands where there were moderate levels of starvation and similar experiences of war, but not quite the extreme caloric deprivation experienced in the worst affected areas. Differential associations were still seen when using these cohorts as a control group, suggesting stress and the experience of wartime are unlikely to account for the results (Brown and Susser, 2008).

## Conclusion

The approaches described in this review represent a number of different ways in which study design and broad analytical methods can be used in order to allow for stronger causal inferences than is provided by conventional statistical adjustments. Negative control designs identify an exposure or outcome where no association is predicted, but which share a similar confounding structure with the main association of interest. This can help rule out residual confounding as an explanation for the association of interest. Cross-contextual studies compare associations between two populations where underlying confounding structures are likely to be very different, which would lessen the likelihood of associations being due to confounding. Instrumental variable analyses identify an unconfounded proxy for the exposure of interest, and assess the association between that and the outcome, to remove the effect of unmeasured confounding. A specific version of this, Mendelian randomisation, utilizes genetic variants as the proxy variables which can also rule out reverse causation, as genes are determined at conception. Family-based studies use shared genetic and environmental characteristics to generate highly matched case-control studies.

While none of these techniques represents a panacea, each having its own strengths and weakness, they can be used in conjunction with each other to provide an overall evaluation of the support for putative causal associations seen in observational data. Combining these different designs in a single report assessing one research question from a variety of angles can be particularly effective. For example, Brion and colleagues (Brion et al., 2010b) combined a cross-contextual and negative control design to assess associations between maternal smoking and child psychological problems. They found that maternal smoking during pregnancy was associated with greater offspring externalizing and peer problems in cohorts in Brazil and United Kingdom, despite smoking during pregnancy being differentially socially patterned between the two countries. The authors also showed that associations between maternal smoking and offspring conduct problems were stronger than those between paternal smoking and the same problems (although statistical evidence was weak in one cohort). By combining these study designs, the findings become much more compelling than they would alone.

These different approaches that use study design to leverage stronger causal inference each rely on specific assumptions, which may not be valid. Critically, however, they rely on different assumptions. The triangulation of evidence from these different methods is therefore a powerful tool, and arguably a much more reliable approach to causal inference than statistical adjustment for imprecisely measured confounders, which are likely to constitute only some of the confounding factors that plague naïve observational epidemiology. Many methods are particularly well suited to the study of the Developmental Origins of Health and Disease, and a number of examples exist of these methods being applied to better understand the causal effects of intrauterine exposures such as tobacco and alcohol use on offspring developmental outcomes. The tools necessary to implement these methods are becoming increasingly widely available. Access to datasets from large cohort studies across different countries is increasing, and a growing number of genetic variants

associated with exposures of interest such as tobacco and alcohol use are being identified via genome wide association studies. The potential for the application of these methods is therefore growing rapidly, offering great promise for future Developmental Origins of Health and Disease research. A few key considerations can ensure the robust triangulation of evidence, such as ensuring that variables across studies are meaningfully and harmoniously coded and scaled, to allow direct comparison across designs. Consideration of the magnitude of effect of a hypothesised dose-response relationship across different study durations can provide stronger evidence in support of causation. For example exposure differences in RCTs are likely to be of much shorter duration than cohort studies, and in Mendelian randomisation studies are likely to be longer than either (as they are present from conception); therefore the magnitude of the observed effect size would be expected to differ across these studies if associations were causal and showed dose-response. Critical periods should also be considered when triangulating the findings from different study designs, which is particularly relevant for research on the Developmental Origins of Health and Disease. It may be that a risk factor only has an effect on an outcome during a specific period of pregnancy, and if different studies measure variables at slightly different times, this could be the reason for inconsistent results, rather than a lack of a causal association. Finally, there may be multiple hypotheses to explain observed associations, and therefore applying principles of inference, and considering possible sources of bias, will be important when attempting to triangulate results across different designs (Richmond et al., 2014).

Acknowledgements

We would like to thank Neil Davies for helpful comments on an earlier draft of this review, and Rebecca Richmond and Gemma Sharp for initial comments. All authors work within the MRC Integrative Epidemiology Unit, supported by the Medical Research Council (MC_UU_12013/1, MC_UU_12013/6) and the University of Bristol. SHG and MRM are members of the UK Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

References


ANGRIST, J. D. & PISCHKE, J.-S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton, Princeton University Press.

ATTERMANN, J., OBEL, C., BILENBERG, N., NORDENBAEK, C. M., SKYTTHE, A. & OLSEN, J. 2012. Traits of ADHD and autism in girls with a twin brother: a Mendelian randomization study. *Eur Child Adolesc Psychiatry,* 21**,** 503-9.

BARKER, D. J. 1995. Fetal origins of coronary heart disease. *BMJ,* 311**,** 171-4.

BARKER, D. J. 1997. Intrauterine programming of coronary heart disease and stroke. *Acta Paediatr Suppl,* 423**,** 178-82; discussion 183.

BARKER, D. J. & MARTYN, C. N. 1997. The fetal origins of hypertension. *Adv Nephrol Necker Hosp,* 26**,** 65-72.

BARKER, D. J. & OSMOND, C. 1986. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet,* 1**,** 1077-81.

BOWDEN, J., DAVEY SMITH, G. & BURGESS, S. 2015. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*.

BRION, M. J., LAWLOR, D. A., MATIJASEVICH, A., HORTA, B., ANSELMI, L., ARAUJO, C. L., MENEZES, A. M., VICTORA, C. G. & SMITH, G. D. 2011. What are the causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts. *Int J Epidemiol,* 40**,** 670-80.

BRION, M. J., LEARY, S. D., SMITH, G. D. & NESS, A. R. 2007. Similar associations of parental prenatal smoking suggest child blood pressure is not influenced by intrauterine effects. *Hypertension,* 49**,** 1422-8.

BRION, M. J., NESS, A. R., ROGERS, I., EMMETT, P., CRIBB, V., DAVEY SMITH, G. & LAWLOR, D. A. 2010a. Maternal macronutrient and energy intakes in pregnancy and offspring intake at 10 y: exploring parental comparisons and prenatal effects. *Am J Clin Nutr,* 91**,** 748-56.

BRION, M. J., VICTORA, C., MATIJASEVICH, A., HORTA, B., ANSELMI, L., STEER, C., MENEZES, A. M., LAWLOR, D. A. & DAVEY SMITH, G. 2010b. Maternal smoking and child psychological problems: disentangling causal and noncausal effects. *Pediatrics,* 126**,** e57-65.

BROWN, A. S. & SUSSER, E. S. 1997. Sex differences in prevalence of congenital neural defects after periconceptional famine exposure. *Epidemiology,* 8**,** 55-8.

BROWN, A. S. & SUSSER, E. S. 2008. Prenatal nutritional deficiency and risk of adult schizophrenia. *Schizophr Bull,* 34**,** 1054-63.

CLASS, Q. A., RICKERT, M. E., LARSSON, H., LICHTENSTEIN, P. & D'ONOFRIO, B. M. 2014. Fetal growth and psychiatric and socioeconomic problems: population-based sibling comparison. *Br J Psychiatry,* 205**,** 355-61.

COLE, S. R., PLATT, R. W., SCHISTERMAN, E. F., CHU, H., WESTREICH, D., RICHARDSON, D. & POOLE, C. 2010. Illustrating bias due to conditioning on a collider. *Int J Epidemiol,* 39**,** 417-20.

D'ONOFRIO, B. M., CLASS, Q. A., LAHEY, B. B. & LARSSON, H. 2014. Testing the Developmental Origins of Health and Disease Hypothesis for Psychopathology Using Family-Based Quasi-Experimental Designs. *Child Dev Perspect,* 8**,** 151-157.

DAVEY SMITH, G. 2008. Assessing intrauterine influences on offspring health outcomes: can epidemiological studies yield robust findings? *Basic Clin Pharmacol Toxicol,* 102**,** 245-56.

DAVEY SMITH, G. 2010. Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene x Environment Interactions. *Perspectives on Psychological Science,* 5**,** 527-545.

DAVEY SMITH, G. 2012. Negative control exposures in epidemiologic studies. *Epidemiology,* 23**,** 350-1; author reply 351-2.

DAVEY SMITH, G. & EBRAHIM, S. 2001. Epidemiology--is it time to call it a day? *Int J Epidemiol,* 30**,** 1-11.
DAVEY SMITH, G. & EBRAHIM, S. 2002. Data dredging, bias, or confounding. *BMJ,* 325**,** 1437-8.
DAVEY SMITH, G. & EBRAHIM, S. 2003. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol,* 32**,** 1-22.
DAVEY SMITH, G. & HEMANI, G. 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet,* 23**,** R89-98.
DAVEY SMITH, G., PHILLIPS, A. N. & NEATON, J. D. 1992. Smoking as "independent" risk factor for suicide: illustration of an artifact from observational epidemiology? *Lancet,* 340**,** 709-12.
DI FORTI, M., MARCONI, A., CARRA, E., FRAIETTA, S., TROTTA, A., BONOMO, M., BIANCONI, F., GARDNER-SOOD, P., O'CONNOR, J., RUSSO, M., STILO, S., REIS MARQUES, T., MONDELLI, V., DAZZAN, P., PARIANTE, C., DAVID, A. S., GAUGHRAN, F., ATAKAN, Z., IYEGBE, C., POWELL, J., MORGAN, C., LYNSKEY, M. & MURRAY, R. M. 2015. Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study. *Lancet Psychiatry*.
DINARDO, J. E. & PISCHKE, J. S. 1997. The returns to computer use revisited: Have pencils changed the wage structure too? *Quarterly Journal of Economics,* 112**,** 291-303.
ENOMOTO, N., TAKASE, S., YASUHARA, M. & TAKADA, A. 1991. Acetaldehyde metabolism in different aldehyde dehydrogenase-2 genotypes. *Alcohol Clin Exp Res,* 15**,** 141-4.
ERIKSSON, J. G., SANDBOGE, S., SALONEN, M., KAJANTIE, E. & OSMOND, C. 2015. Maternal weight in pregnancy and offspring body composition in late adulthood: Findings from the Helsinki Birth Cohort Study (HBCS). *Ann Med***,** 1-6.
ERYIGIT MADZWAMUSE, S., BAUMANN, N., JAEKEL, J., BARTMANN, P. & WOLKE, D. 2014. Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. *J Child Psychol Psychiatry*.
FALL, C. H., OSMOND, C., BARKER, D. J., CLARK, P. M., HALES, C. N., STIRLING, Y. & MEADE, T. W. 1995. Fetal and infant growth and cardiovascular risk factors in women. *BMJ,* 310**,** 428-32.
FEWELL, Z., DAVEY SMITH, G. & STERNE, J. A. 2007. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol,* 166**,** 646-55.
FREATHY, R. M., RING, S. M., SHIELDS, B., GALOBARDES, B., KNIGHT, B., WEEDON, M. N., SMITH, G. D., FRAYLING, T. M. & HATTERSLEY, A. T. 2009. A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum Mol Genet,* 18**,** 2922-7.
GAGE, S. H., MUNAFO, M. R., MACLEOD, J., HICKMAN, M. & DAVEY SMITH, G. 2015. Cannabis and psychosis. *Lancet Psychiatry,* 2**,** 380.
GAYSINA, D., FERGUSSON, D. M., LEVE, L. D., HORWOOD, J., REISS, D., SHAW, D. S., ELAM, K. K., NATSUAKI, M. N., NEIDERHISER, J. M. & HAROLD, G. T. 2013. Maternal smoking during pregnancy and offspring conduct problems: evidence from 3 independent genetically sensitive research designs. *JAMA Psychiatry,* 70**,** 956-63.
HALES, C. N. & BARKER, D. J. 1992. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia,* 35**,** 595-601.
HOEK, H. W., BROWN, A. S. & SUSSER, E. 1998. The Dutch famine and schizophrenia spectrum disorders. *Soc Psychiatry Psychiatr Epidemiol,* 33**,** 373-9.

KELLY, Y. J., SACKER, A., GRAY, R., KELLY, J., WOLKE, D., HEAD, J. & QUIGLEY, M. A. 2012. Light drinking during pregnancy: still no increased risk for socioemotional difficulties or cognitive deficits at 5 years of age? *J Epidemiol Community Health,* 66**,** 41-8.

KESKITALO, K., BROMS, U., HELIOVAARA, M., RIPATTI, S., SURAKKA, I., PEROLA, M., PITKANIEMI, J., PELTONEN, L., AROMAA, A. & KAPRIO, J. 2009. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. *Hum Mol Genet,* 18**,** 4007-12.

KING, L., CARPENTIER, C. & GRIFFITHS, P. 2004. *EMCDDA Insights: An overview of cannabis potency in Europe,* Luxembourg, European Monitoring Centre for Drugs and Drug Addiction.

KRAMER, M. S., MATUSH, L., VANILOVICH, I., PLATT, R. W., BOGDANOVICH, N., SEVKOVSKAYA, Z., DZIKOVICH, I., SHISHKO, G., COLLET, J. P., MARTIN, R. M., DAVEY SMITH, G., GILLMAN, M. W., CHALMERS, B., HODNETT, E., SHAPIRO, S. & GROUP, P. S. 2007. Effects of prolonged and exclusive breastfeeding on child height, weight, adiposity, and blood pressure at age 6.5 y: evidence from a large randomized trial. *Am J Clin Nutr,* 86**,** 1717-21.

KUH, D. & DAVEY SMITH, G. 2004. The life course and adult chronic disease: an historical perspective with particular reference to coronary heart disease. *In:* KUH, D. & BEN-SHLOMO, Y. (eds.) *A Life Course Approach to Chronic Disease Epidemiology (2nd edition).* Oxford: Oxford University Press.

LAWLOR, D. A., SMITH, G. D. & EBRAHIM, S. 2004. Socioeconomic position and hormone replacement therapy use: explaining the discrepancy in evidence from observational and randomized controlled trials. *Am J Public Health,* 94**,** 2149-54.

LAWLOR, D. A., TIMPSON, N. J., HARBORD, R. M., LEARY, S., NESS, A., MCCARTHY, M. I., FRAYLING, T. M., HATTERSLEY, A. T. & SMITH, G. D. 2008. Exploring the developmental overnutrition hypothesis using parental-offspring associations and FTO as an instrumental variable. *PLoS Med,* 5**,** e33.

LIANG, W. & CHIKRITZHS, T. 2013. Observational research on alcohol use and chronic disease outcome: new approaches to counter biases. *ScientificWorldJournal,* 2013**,** 860915.

LIPSITCH, M., TCHETGEN TCHETGEN, E. & COHEN, T. 2010. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology,* 21**,** 383-8.

LORENZ, K. 1935. Der Kumpan in der Umwelt des Vogels. *Journal of Ornithology,* 83**,** 137-213.

MARTIN, R. M., GOODALL, S. H., GUNNELL, D. & DAVEY SMITH, G. 2007. Breast feeding in infancy and social mobility: 60-year follow-up of the Boyd Orr cohort. *Arch Dis Child,* 92**,** 317-21.

MEHMEDIC, Z., CHANDRA, S., SLADE, D., DENHAM, H., FOSTER, S., PATEL, A. S., ROSS, S. A., KHAN, I. A. & ELSOHLY, M. A. 2010. Potency trends of Delta9-THC and other cannabinoids in confiscated cannabis preparations from 1993 to 2008. *J Forensic Sci,* 55**,** 1209-17.

MUNAFO, M. R., TIMOFEEVA, M. N., MORRIS, R. W., PRIETO-MERINO, D., SATTAR, N., BRENNAN, P., JOHNSTONE, E. C., RELTON, C., JOHNSON, P. C., WALTHER, D., WHINCUP, P. H., CASAS, J. P., UHL, G. R., VINEIS, P., PADMANABHAN, S., JEFFERIS, B. J., AMUZU, A., RIBOLI, E., UPTON, M. N., AVEYARD, P., EBRAHIM, S., HINGORANI, A. D., WATT, G., PALMER, T. M., TIMPSON, N. J. & DAVEY SMITH, G. 2012. Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *J Natl Cancer Inst,* 104**,** 740-8.

NILSSON, J. P. 2014. Alcohol Availability, Prenatal Conditions, and Long-term Economic Outcomes. *Working Paper, Institute for International Economic Studies*.
OOSTERBEEK, H. 1997. Returns from computer use: A simple test on the productivity interpretation. *Economics Letters,* 55**,** 273-277.
PATEL, R., OKEN, E., BOGDANOVICH, N., MATUSH, L., SEVKOVSKAYA, Z., CHALMERS, B., HODNETT, E. D., VILCHUCK, K., KRAMER, M. S. & MARTIN, R. M. 2014. Cohort profile: The promotion of breastfeeding intervention trial (PROBIT). *Int J Epidemiol,* 43**,** 679-90.
PETITTI, D. B., PERLMAN, J. A. & SIDNEY, S. 1986. Postmenopausal Estrogen Use and Heart-Disease. *New England Journal of Medicine,* 315**,** 131-132.
PETITTI, D. B., PERLMAN, J. A. & SIDNEY, S. 1987. Noncontraceptive estrogens and mortality: long-term follow-up of women in the Walnut Creek Study. *Obstet Gynecol,* 70**,** 289-93.
PITKIN, R. M. 2007. Folate and neural tube defects. *Am J Clin Nutr,* 85**,** 285S-288S.
RICE, F., HAROLD, G. T., BOIVIN, J., HAY, D. F., VAN DEN BREE, M. & THAPAR, A. 2009. Disentangling prenatal and inherited influences in humans with an experimental design. *Proc Natl Acad Sci U S A,* 106**,** 2464-7.
RICHMOND, R. C., AL-AMIN, A., SMITH, G. D. & RELTON, C. L. 2014. Approaches for drawing causal inferences from epidemiological birth cohorts: A review. *Early Human Development,* 90**,** 769-780.
ROTH, C., MAGNUS, P., SCHJOLBERG, S., STOLTENBERG, C., SUREN, P., MCKEAGUE, I. W., DAVEY SMITH, G., REICHBORN-KJENNERUD, T. & SUSSER, E. 2011. Folic acid supplements in pregnancy and severe language delay in children. *JAMA,* 306**,** 1566-73.
ROTHMAN, K., GREENLAND, S. & LASH, T. L. 2008. *Modern Epidemiology,* Philidelphia, Lippincott Williams & Wilkins.
RUBINO, T. & PAROLARO, D. 2008. Long lasting consequences of cannabis exposure in adolescence. *Mol Cell Endocrinol,* 286**,** S108-13.
SONG, S., WANG, W. & HU, P. 2009. Famine, death, and madness: schizophrenia in early adulthood after prenatal exposure to the Chinese Great Leap Forward Famine. *Soc Sci Med,* 68**,** 1315-21.
SUREN, P., ROTH, C., BRESNAHAN, M., HAUGEN, M., HORNIG, M., HIRTZ, D., LIE, K. K., LIPKIN, W. I., MAGNUS, P., REICHBORN-KJENNERUD, T., SCHJOLBERG, S., DAVEY SMITH, G., OYEN, A. S., SUSSER, E. & STOLTENBERG, C. 2013. Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. *JAMA,* 309**,** 570-7.
TAYLOR, A. E., DAVEY SMITH, G., BARES, C. B., EDWARDS, A. C. & MUNAFO, M. R. 2014. Partner smoking and maternal cotinine during pregnancy: implications for negative control methods. *Drug Alcohol Depend,* 139**,** 159-63.
THAPAR, A., RICE, F., HAY, D., BOIVIN, J., LANGLEY, K., VAN DEN BREE, M., RUTTER, M. & HAROLD, G. 2009. Prenatal smoking might not cause attention-deficit/hyperactivity disorder: evidence from a novel design. *Biol Psychiatry,* 66**,** 722-7.
THORGEIRSSON, T. E., GELLER, F., SULEM, P., RAFNAR, T., WISTE, A., MAGNUSSON, K. P., MANOLESCU, A., THORLEIFSSON, G., STEFANSSON, H., INGASON, A., STACEY, S. N., BERGTHORSSON, J. T., THORLACIUS, S., GUDMUNDSSON, J., JONSSON, T., JAKOBSDOTTIR, M., SAEMUNDSDOTTIR, J., OLAFSDOTTIR, O., GUDMUNDSSON, L. J., BJORNSDOTTIR, G., KRISTJANSSON, K., SKULADOTTIR, H., ISAKSSON, H. J., GUDBJARTSSON, T., JONES, G. T., MUELLER, T., GOTTSATER, A., FLEX, A., ABEN, K. K., DE VEGT, F., MULDERS, P. F., ISLA, D., VIDAL, M. J., ASIN, L., SAEZ, B., MURILLO, L., BLONDAL, T., KOLBEINSSON, H., STEFANSSON, J. G., HANSDOTTIR, I., RUNARSDOTTIR, V., POLA, R.,

LINDBLAD, B., VAN RIJ, A. M., DIEPLINGER, B., HALTMAYER, M., MAYORDOMO, J. I., KIEMENEY, L. A., MATTHIASSON, S. E., OSKARSSON, H., TYRFINGSSON, T., GUDBJARTSSON, D. F., GULCHER, J. R., JONSSON, S., THORSTEINSDOTTIR, U., KONG, A. & STEFANSSON, K. 2008. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature,* 452**,** 638-42.

TREZZA, V., CUOMO, V. & VANDERSCHUREN, L. J. 2008. Cannabis and the developing brain: insights from behavior. *Eur J Pharmacol,* 585**,** 441-52.

TYRRELL, J., HUIKARI, V., CHRISTIE, J. T., CAVADINO, A., BAKKER, R., BRION, M. J., GELLER, F., PATERNOSTER, L., MYHRE, R., POTTER, C., JOHNSON, P. C., EBRAHIM, S., FEENSTRA, B., HARTIKAINEN, A. L., HATTERSLEY, A. T., HOFMAN, A., KAAKINEN, M., LOWE, L. P., MAGNUS, P., MCCONNACHIE, A., MELBYE, M., NG, J. W., NOHR, E. A., POWER, C., RING, S. M., SEBERT, S. P., SENGPIEL, V., TAAL, H. R., WATT, G. C., SATTAR, N., RELTON, C. L., JACOBSSON, B., FRAYLING, T. M., SORENSEN, T. I., MURRAY, J. C., LAWLOR, D. A., PENNELL, C. E., JADDOE, V. W., HYPPONEN, E., LOWE, W. L., JR., JARVELIN, M. R., DAVEY SMITH, G., FREATHY, R. M. & EARLY GROWTH GENETICS, C. 2012. Genetic variation in the 15q25 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) interacts with maternal self-reported smoking status during pregnancy to influence birth weight. *Hum Mol Genet,* 21**,** 5344-58.

VAN LIESHOUT, R. J. & BOYLAN, K. 2010. Increased depressive symptoms in female but not male adolescents born at low birth weight in the offspring of a national cohort. *Can J Psychiatry,* 55**,** 422-30.

VICTORA, C. G., HORTA, B. L., LORET DE MOLA, C., QUEVEDO, L., PINHEIRO, R. T., GIGANTE, D. P., GONCALVES, H. & BARROS, F. C. 2015. Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a prospective birth cohort study from Brazil. *Lancet Glob Health,* 3**,** e199-205.

WARE, J. J., VAN DEN BREE, M. & MUNAFO, M. R. 2012. From Men to Mice: CHRNA5/CHRNA3, Smoking Behavior and Disease. *Nicotine Tob Res*.

WARE, J. J., VAN DEN BREE, M. B. & MUNAFO, M. R. 2011. Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: a meta-analysis. *Nicotine Tob Res,* 13**,** 1167-75.

WATSON, J. B. & RAYNER, R. 1920. Conditioned emotional reactions. *Journal of Experimental Psychology,* 3**,** 1-14.

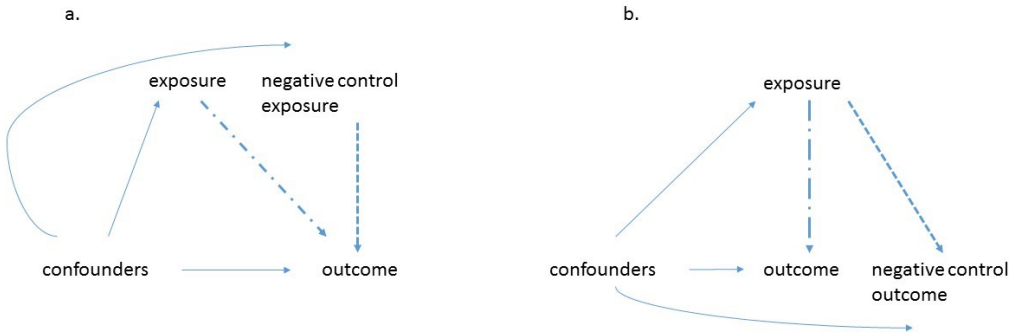
ZAMMIT, S., ALLEBECK, P., ANDREASSON, S., LUNDBERG, I. & LEWIS, G. 2002. Self reported cannabis use as a risk factor for schizophrenia in Swedish conscripts of 1969: historical cohort study. *BMJ,* 325**,** 1199.

ZUCCOLO, L., LEWIS, S. J., SMITH, G. D., SAYAL, K., DRAPER, E. S., FRASER, R., BARROW, M., ALATI, R., RING, S., MACLEOD, J., GOLDING, J., HERON, J. & GRAY, R. 2013. Prenatal alcohol exposure and offspring cognition and school performance. A 'Mendelian randomization' natural experiment. *Int J Epidemiol,* 42**,** 1358-70.

**Table 1. Description of the methodologies reviewed.**

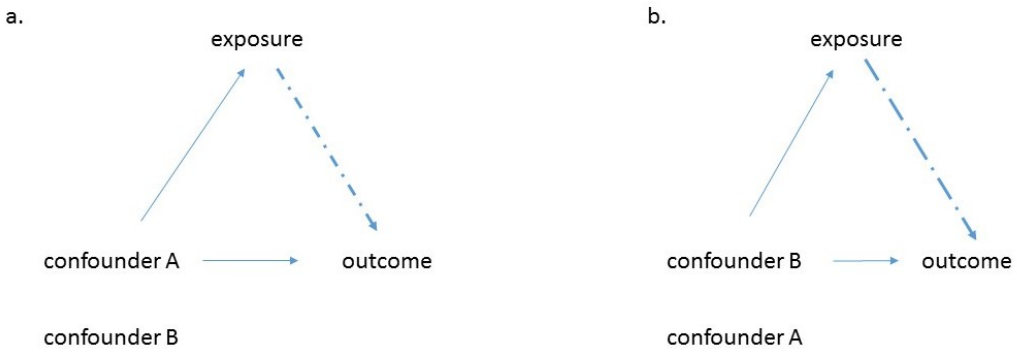
| Technique | Summary |
|---|---|
| Negative control | Exposures or outcomes are identified with similar confounding but no plausible biological connection to ascertain whether associations are likely to be causal or due to confounding. |
| Cross-contextual | Two populations with differing confounding structures are sampled and associations compared between them. |
| Instrumental variable analysis | Unconfounded proxies are found for exposures of interest (e. g. genetic variants in Mendelian randomization). |
| Family studies | Related pairs of individuals are compared, where assumptions can be made about shared genetic and environmental factors. |

**Figure 1. Schematic representations showing a) negative control exposure and b) negative control outcome.**



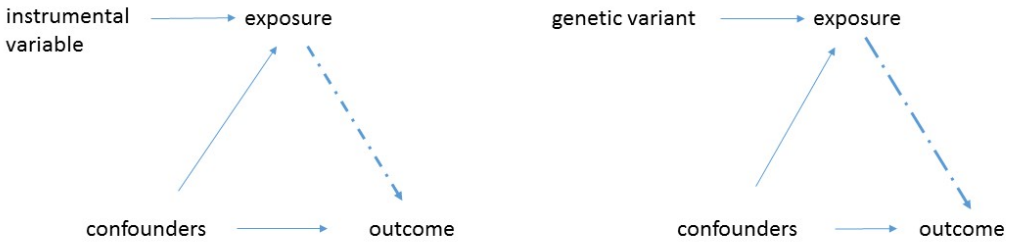

The dotted and dashed line represents the association of interest. Confounding is the same for the exposure or outcome and its' negative control. However, there is no causal association between: a) the negative control exposure and the outcome of interest, or b) the exposure of interest and negative control outcome. The dashed line represents the negative control analysis.

**Figure 2. Schematic representations showing a cross-contextual design.**



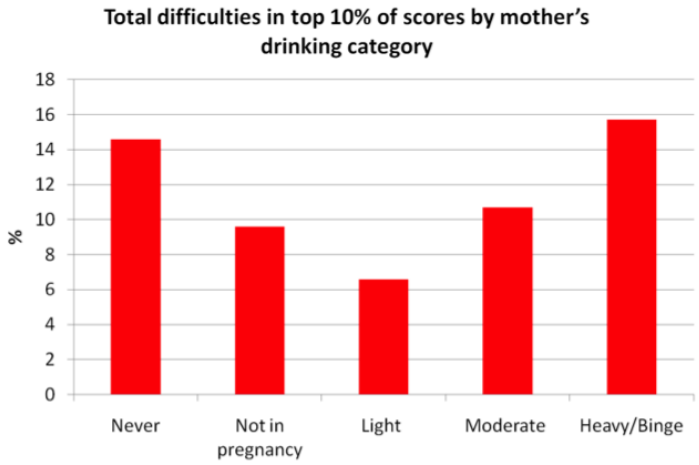

The exposure and outcome should be equivalent across the different contexts, but the confounding structure should not. Here, confounder A impacts upon the relationship in context a) but not context b). The reverse is true for confounder B.

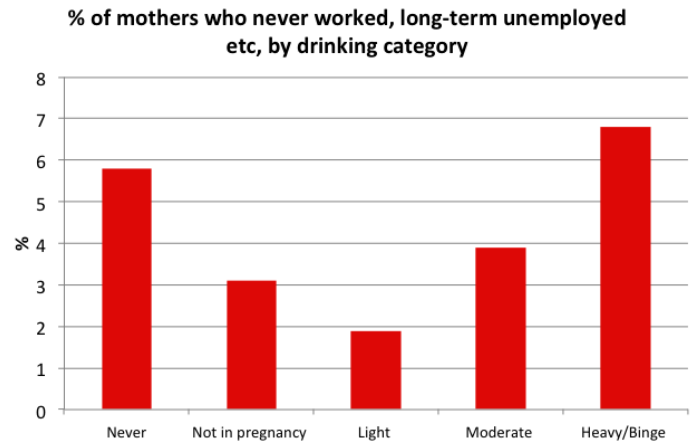
**Figure 3. Schematic representation showing a) an instrumental variables analysis and b) a Mendelian randomisation analysis.**

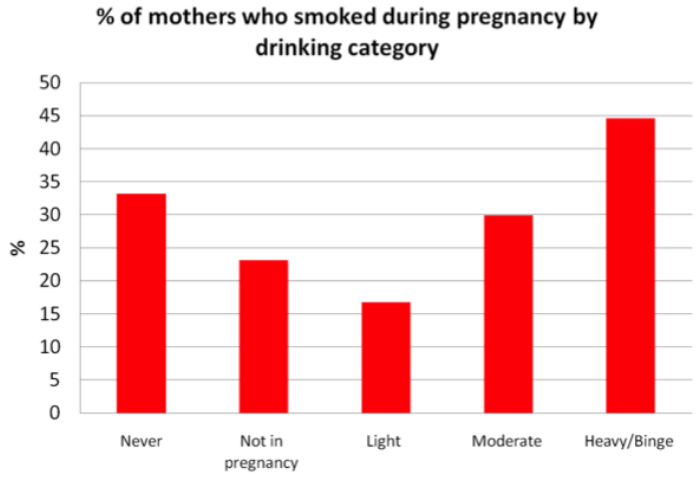



The instrument or genetic variant is associated with the exposure of interest, but not with the confounding variables associated with the exposure and outcome. It is only associated with the outcome via its association with the exposure of interest.

**Figure 4. Association of maternal alcohol consumption in pregnancy with offspring behavioural difficulties, maternal socioeconomic position, and maternal smoking.**




Adapted from Kelly et al. (2012).

**TEXT BOX: IVF variation.**

A novel design has recently been used to attempt to disentangle prenatal from inherited effects, using pregnancies resulting from vitro fertilisation (IVF). In some instances IVF pregnancies will use embryos harvested from the woman who will carry the child, but in other cases an embryo from a different woman will be implanted, so the “mother” who will carry the child will not be biologically related to it. If a particular outcome is associated with an exposure occurring during pregnancy regardless of the biological relatedness of the mother and offspring, it suggests that the association is likely to be due to the intrauterine environment. However, if the association is only seen where the mother is biologically related to the offspring, this indicates that genetic confounding might be driving the association seen. The technique has been used to assess the impact of smoking during pregnancy. Rice and colleagues (Rice et al., 2009) found that smoking was associated with reduced offspring birthweight regardless of whether the mother was biologically related to the offspring or not. This pattern of findings was also shown in a different sample (Thapar et al., 2009). However, the association between smoking and offspring antisocial behaviour was dependent on inherited factors, as it was only seen in biologically related pairs, and not where the offspring . A later study, combining data using this study design with two studies assessing adoption, found converging evidence for an intrauterine effect of smoking on offspring conduct problems (Gaysina et al., 2013).

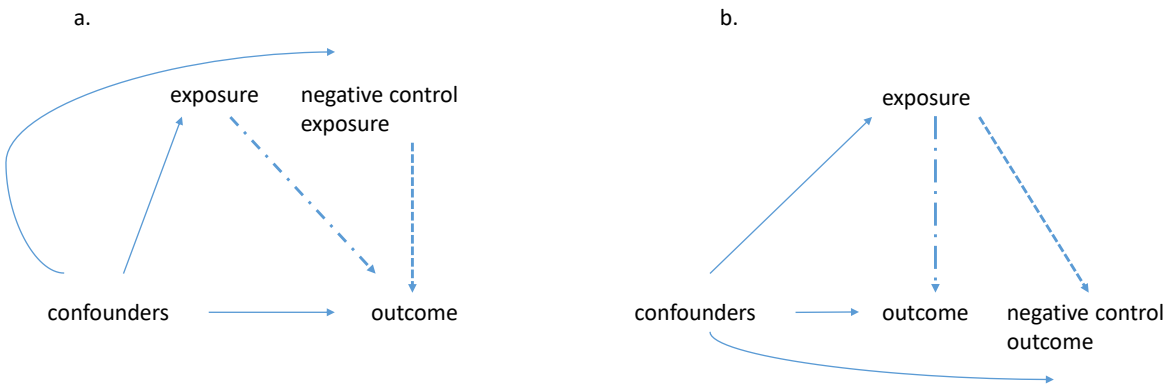



**Figure 1. Schematic representations showing a) negative control exposure and b) negative control outcome.** The dotted and dashed line represents the association of interest. Confounding is the same for the exposure or outcome and its' negative control. However, there is no causal association between: (a) the negative control exposure and the outcome of interest, or (b) the exposure of interest and negative control outcome. The dashed line represents the negative control analysis.

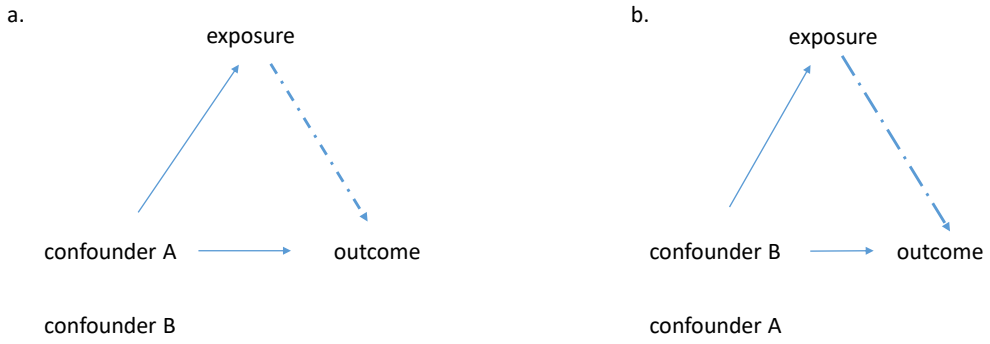



**Figure 2. Schematic representations showing a cross-contextual design.** The exposure and outcome should be equivalent across the different contexts, but the confounding structure should not. Here, confounder A impacts upon the relationship in context (a) but not context (b). The reverse is true for confounder B.

Figure 3

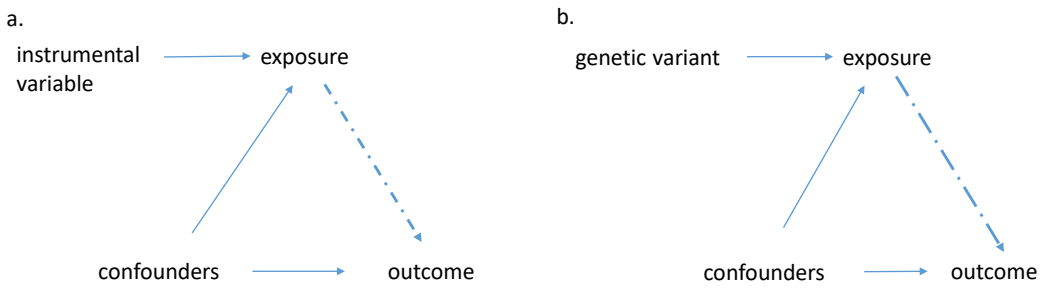



**Figure 3. Schematic representation showing (a) an instrumental variables analysis and (b) a Mendelian randomization analysis.** The instrument or genetic variant is associated with the exposure of interest, but not with the confounding variables associated with the exposure and outcome. It is only associated with the outcome via its association with the exposure of interest.