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

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

[1]MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol BS8 2BN, United Kingdom; email: kz.davey-smith@bristol.ac.uk

[2]UK Center for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol BS8 1TU, United Kingdom

Annu. Rev. Psychol. 2016. 67:567–85

First published online as a Review in Advance on October 6, 2015

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi: 10.1146/annurev-psych-122414-033352

Keywords

DOHaD, causal inference, instrumental variable, negative control, cross-contextual comparison, twin study

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 methods 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 behavioral, psychiatric, and psychological domains, are considered, and the considerable potential for expanding the use of these methods is outlined.

Contents

INTRODUCTION

The developmental origins of health and disease (DOHaD) hypothesis proposes that the environment individuals experience 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 DOHaD (Kuh & 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, Barker & Osmond 1986, Fall et al. 1995). 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 & Rayner 1920), theories conceptually similar to the DOHaD hypothesis have played a key role in psychological research over the preceding century. 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.

A substantial body of evidence supports 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 & Martyn 1997), and type 2 diabetes (Hales & Barker 1992). These studies were originally considered under the umbrella term of the fetal origins of adult disease (FOAD) because 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 postnatal period and possible preconceptual and intergenerational influences (Davey Smith 2012a). 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 & Boylan 2010), and intelligence (Eryigit Madzwamuse et al. 2015). However, studies to date have for the most part been conducted using observational data; therefore, although they provide suggestive evidence that these developmental influences affect later outcomes, they are limited in terms of providing strong enough evidence for a causal interpretation to 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 are 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 randomized controlled trials (RCTs), in which individuals are randomly assigned to either an exposure condition or a control condition and then 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 either to give or to 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 possibilities aside from a causal association between the two. These possibilities include confounding, reverse causation, and various biases that could distort the underlying association (Davey Smith & Ebrahim 2002).

Confounding. If other differences exist 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 (Phillips & Davey Smith 1992). Unmeasured confounding factors and measurement error (due to either technical issues or 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 (Davey Smith & Phillips 1992, 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 that 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 nondrinkers stop drinking due to ill health; the drinking behavior is a consequence of the increased risk, rather than the other way round (Liang & 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 (Ebrahim

& Davey Smith 2013). This is a form of collider bias (Cole et al. 2010)—a family of biases that can distort observational estimates of exposure effects—that has perhaps been underappreciated 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 minimizing them. However, statistical analyses necessarily require assumptions to be made about the data, and these may be (and indeed probably usually are) unjustified. Moreover, statistical adjustments can lead to overconfidence in the robustness of findings—the commonly used term that factors have been statistically controlled for gives a sense of this—and result in the literature containing many associations that are overinterpreted in terms of causal evidence (Davey Smith & Ebrahim 2001). As with residual confounding, attempted statistical adjustment to account for potential biases has serious limitations.

In this review, we discuss improving causal inference through alternative approaches to conventional statistical adjustment methods. Ways of assembling and analyzing data can strengthen such inference, and triangulating evidence from multiple independent sources can provide more reliable evidence for causation than would a single approach. In lieu of experimental designs, which are typically not ethical or practical, a far stronger foundation for this literature can be built through identifying study designs that 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 (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 rely 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** summarizes these methods.

Table 1 Description of the methodologies reviewed

Technique	Summary
Negative control	Exposures or outcomes with similar confounding but no plausible biological connection are identified 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 are compared between them.
Instrumental variable analysis	Unconfounded proxies are found for exposures of interest (e.g., genetic variants in Mendelian randomization).
Family studies	Assumptions are made about shared genetic and environmental factors in comparisons of related pairs of individuals.

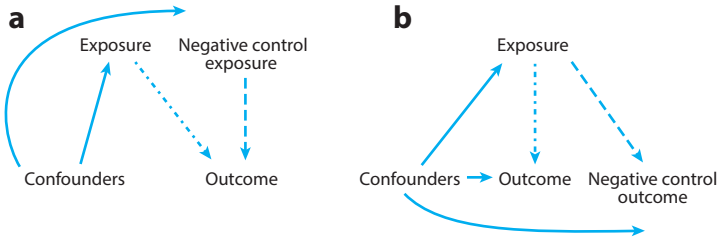

Figure 1

Schematic representations of (*a*) negative control exposure and (*b*) negative control outcome. 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 the negative control outcome. The dashed line represents the negative control analysis, and the dotted-and-dashed line represents the association under interrogation.

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 & 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 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 the negative control outcome. If the association seen between the exposure and outcome is of a larger magnitude than the association between the exposure and the negative control, this would contribute positively to an evaluation of the strength of evidence for a causal association between the exposure and outcome of interest. If, however, the association seen between the two is due to confounding, then a similar association is likely to be seen in both the analysis of interest and the negative control analysis, in which there is no biologically plausible mechanism for causation (Davey Smith 2008, 2012b; Lipsitch et al. 2010). **Figure 1** shows an example of a negative control exposure and a negative control outcome design.

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 hypothesized 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 would 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; although apparently plausible causal biological mechanisms exist 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 lack of 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 socioeconomic position or health-adverse maternal behaviors 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 the use of folic acid supplements were similarly socially patterned with respect to potential confounders such as parental characteristics. A robust inverse association was identified between the use of folic acid supplements and the subsequent risk of autism spectrum disorder. However, there was little evidence of an association between the use of fish oil supplements and autism spectrum disorder. The difference between these two results provides evidence that residual confounding from maternal health-related behaviors or social circumstances more generally is not 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 not containing folic acid and later language delay, compared to the baseline of "no supplement use," despite the similar associations with confounding factors shown for the different supplements. 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 simply 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 (Davey Smith 2008). Brion and colleagues found that maternal macronutrient and energy intake during pregnancy predicted later offspring dietary intake, whereas paternal nutrition during the partner's pregnancy did not. They concluded that this provides some evidence that maternal intake during pregnancy could program 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 affects the associations (Brion et al. 2007).

Negative control outcomes. Negative control outcomes use broadly the same principles as negative control exposures. An outcome variable is selected that is unlikely to be caused 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 1980s, 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 assess rates of mortality from accidents, suicide, and homicide in women using HRT compared to those not using it, in whom there is no plausible biological mechanism (Petitti et al. 1986, 1987). 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 nonusers were likely to be due to lifestyle, socioeconomic, behavioral, 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 could be subject to bias. The technique is also inappropriate if there might be a plausible biological mechanism that affects the negative control. For example, paternal smoking during pregnancy could conceivably affect the developing fetus 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 in which the underlying confounding structures are very different, the association provides stronger evidence of causality. It would mean that if the association were due to residual confounding, it would have to be from different sources in the different populations, which is not likely. **Figure 2** illustrates this concept.

One way this design has been utilized is by comparing birth cohorts in different countries. Brion and colleagues (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. Whereas breastfeeding in the United

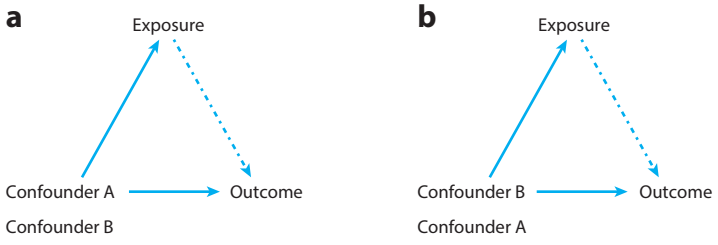

Figure 2

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

Kingdom is associated with higher socioeconomic position, healthier diet, and lower levels of maternal smoking, the same social patterning of breastfeeding behavior did not exist at the time the Brazilian cohort was established. The authors found that in the ALSPAC cohort, breastfeeding was associated with lower offspring body mass index and blood pressure, but 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. Some evidence indicated that the association between breastfeeding and IQ might be causal, and indeed this has been supported by results from an 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 studies in Brazil (Victora et al. 2015).

A second example of cross-contextual comparison is the examination of cohorts in which the patterning of an exposure has changed over time. For example, breastfeeding was not strongly socially patterned in the United Kingdom in the 1920s. Martin and colleagues (2007) compared bottle-fed and breastfed infants in the Boyd-Orr Survey of Diet and Health; the infants were born in the 1920s and 1930s and surveyed between 1937 and 1939 and again between 1997 and 1998. The authors found that breastfeeding in the 1920s was associated with upward social mobility (i.e., 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. Since confounding structures have changed over time, consistency in associations of breastfeeding with offspring health outcomes at different time points increases confidence that these associations are causal in nature.

Conversely, if associations change over time, this suggests that the association may not be causal. Consider one example of this in cannabis research. Some evidence indicates that cannabis use during adolescence could be damaging to later mental health due to changes that occur in the endocannabinoid system during this period of development (Rubino & 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 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), and the size of the point estimates in individual studies has not increased in a systematic way since then. A recent case-control study suggested that skunk cannabis (which is high in THC but has little cannabidiol) is associated with hospitalization for first-episode psychosis, but hash cannabis (with equivalent levels of THC and cannabidiol) is not (Di Forti et al. 2015). Although this could provide evidence in support of an effect of cannabis strength on the 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 cannabis use distribution and are switching to stronger strains as they become available (Gage et al. 2015).

Limitations. When conducting such cross-contextual studies, it is important to ascertain whether relevant differences exist in confounding structure between the two populations being compared. If similar confounders in both contexts 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.

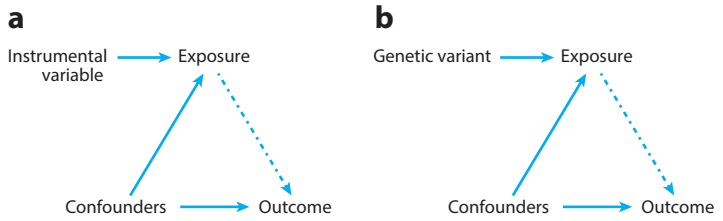

Figure 3

Schematic representations of (*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. The instrument is associated with the outcome only via its association with the exposure of interest.

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 to 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 & Pischke 2009). A diagram of these requirements is shown in **Figure 3*a***.

One study used a short-lived policy change, which had unintended consequences, in an instrumental variable design. In Sweden, a law was introduced that substantially increased access to strong beer by those under 21 in some regions but not other regions. 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; thus, the pregnancies were not due to an increase in unplanned pregnancies resulting from risky sexual behaviors following increased alcohol consumption. The study indicated 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 people born to mothers pregnant in the regions in which policy change occurred but in utero just before or just after this change (Nilsson 2014).

Limitations. The principal limitation of instrumental variable methods is the challenge of identifying 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 because unmeasured confounders may be operating.

Mendelian Randomization

Mendelian randomization is a type of instrumental variable analysis that uses genetic variants as unconfounded proxies (i.e., instruments) for the exposure of interest (Burgess et al. 2015). 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 & Ebrahim 2003). This concept is illustrated in **Figure 3*b***. With regard to developmental outcomes, single nucleotide polymorphisms (SNPs) or genetic risk scores have already been identified via genome-wide association studies for use 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 levels of cotinine, the primary metabolite of nicotine and a more precise biomarker of exposure (Keskitalo et al. 2009, Munafò et al. 2012). The variant has also been shown to associate with lack of ability to give up smoking during pregnancy, which is crucial for investigating developmental outcomes (Freathy et al. 2009). This variant has been used in a number of Mendelian randomization designs, including as a proxy for fetal exposure to cigarette smoke. Tyrrell and colleagues (2012) 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. 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 20-g reduction in offspring birth weight. Conversely, the authors found little evidence of differences in birth weight by genotype in nonsmokers. Given the genotype's lack of association with factors that usually confound observational associations such as age, socioeconomic 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. As previously noted, many observational studies have suggested that the association between alcohol use and many outcomes is J-shaped, indicating that those who drink a small amount have better outcomes than those who do not drink at all. However, drinking behavior is highly socially patterned, so residual confounding could well still be affecting these findings. For example, Kelly and colleagues (2012) reported that low levels of maternal alcohol consumption in pregnancy (one to two drinks per week or per occasion) were associated with reduced behavioral 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 to have been unemployed long term in comparison with 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 adequately captured in the collected data. Zuccolo and colleagues (2013) found evidence that a genotype associated with lower alcohol consumption or abstinence during early pregnancy was associated with offspring academic achievement at age 11, which suggests that alcohol exposure in utero is causally associated with lower offspring educational outcomes. However, these investigators found no strong evidence of an association with childhood IQ at age 8, although the statistical power was lower for this analysis.

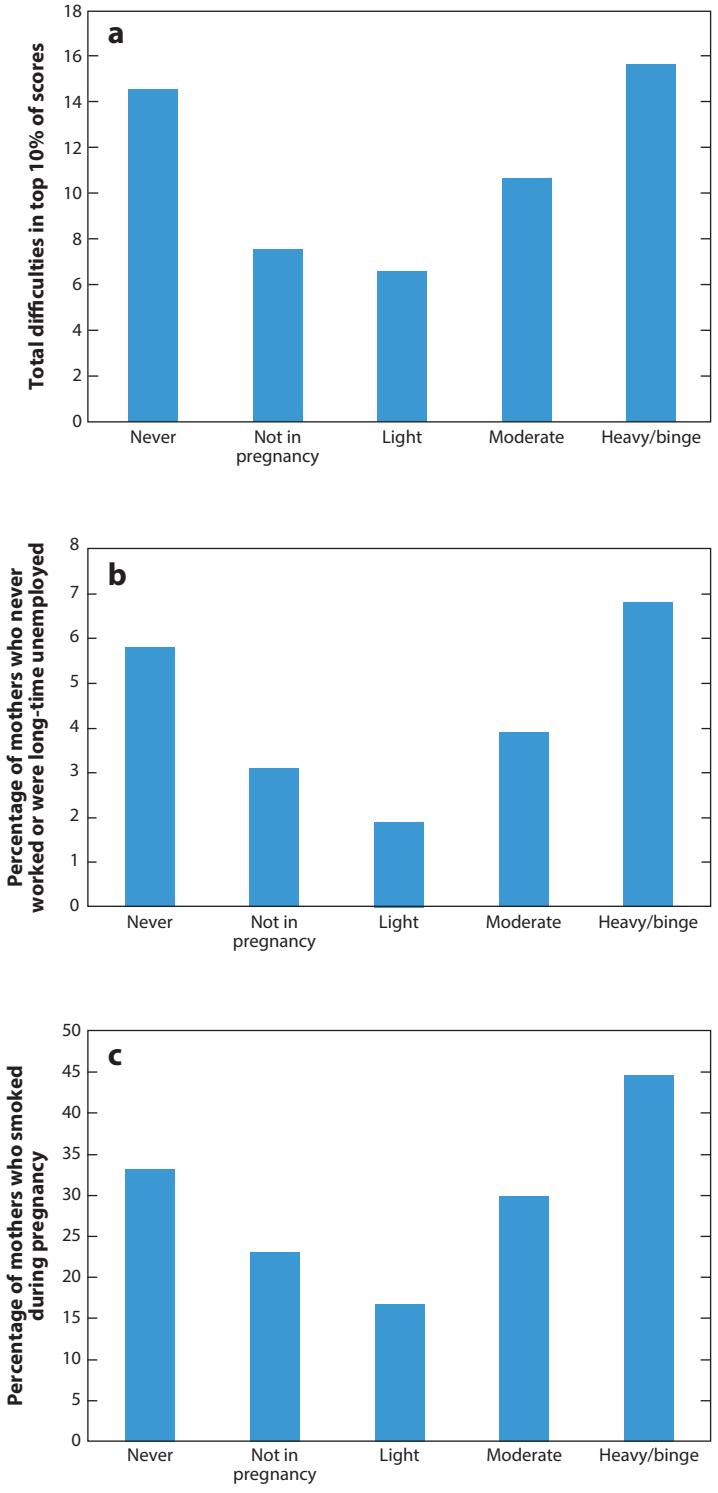

Figure 4

Association of mother's drinking status during pregnancy with (*a*) offspring behavioral difficulties, (*b*) maternal employment history, and (*c*) maternal smoking. Figure based on data from Kelly et al. (2012).

Genetic variants that predict adiposity have been used as a proxy for maternal body mass index in order to ascertain the potential programming effect of prenatal maternal obesity on offspring outcomes. Lawlor and colleagues (2008) found little evidence that maternal genotype predicted offspring fat mass by age 9 to 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 finding suggests that the association between maternal and offspring adiposity may not operate via the prenatal environment.

Limitations. Mendelian randomization is an inappropriate study design in a number of circumstances. Most obviously, it is not possible to use the design if there is no genetic variant yet identified that is robustly associated with the exposure of interest. For example, although cannabis use is known to be heritable, genome-wide association studies have not yet identified any variants robustly associated with cannabis use phenotypes. Given that the associations between variants and exposures of interest are often of modest size, large samples are required for adequate power to undertake such study designs, which means that 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 on 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 biases similar to those seen for pleiotropy. Methods to evaluate and account for such reintroduced confounding for pleiotropy or linkage disequilibrium are discussed elsewhere (Bowden et al. 2015, Davey Smith & 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 to a causal effect of the exposure on the outcome. These issues can be addressed through study designs that utilize ancestrally similar samples and by using principal components analysis to correct for stratification and account for ancestry (Davey Smith & 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 (the latter referring to factors that are the same for both siblings when they grow up together) are less plausible explanations of observed associations (Davey Smith 2008, D'Onofrio et al. 2014). Twin and sibling designs have been used for many years, and a number of different possible designs exist for such studies. The classic twin study design compares monozygotic and dizygotic twin pairs in order to separate additive genetic influence (correlated at 1.0 in monozygotic twins and 0.5 in dizygotic twins), shared environment influence (correlated in both types of twins at 1.0), and nonshared environment influence (not correlated in either type of twins) (Davey Smith 2011).

If pairs of monozygotic twins reared together are discordant for 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 nonshared 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 those with 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 (2014) used a sibling comparison design approach

to disentangle genetic and environmental effects on associations between fetal growth and psychiatric and socioeconomic problems. They 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 his or her substance use, suggesting that residual confounding may have been responsible for associations seen in more traditional cohort designs in which nonrelated individuals were sampled.

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

Limitations. Different limitations exist depending on the type of twin or family study design employed. 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 studies being underpowered. The potential lack of generalizability due to different intrauterine experiences for twin versus singleton births can also limit the impact of some of these study designs, particularly when assessing exposures occurring prenatally.

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. Toward the end of World War II, a Nazi blockade led to a severe food shortage in the Netherlands, where civilians were subjected to rations equivalent to less than 500 calories per day. Pregnancies that occurred during this period represent a rare opportunity to experimentally investigate the impact of severe calorie 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) in comparison with the background rate of such disorders in the Dutch population (Brown & Susser 1997). When the cohort was older, an association with schizophrenia was also assessed, which found the cumulative risk of schizophrenia between ages 24 and 48 years to be double that of unaffected comparison cohorts and of 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 caloric intake data were not available for this cohort, it was still possible to assess the impact of famine during conception on the risk of later schizophrenia. The impacts of severe caloric restriction in this very culturally different cohort were largely similar to those seen in the Netherlands (Song et al. 2009).

Limitations. Such extreme events as famine may have other consequences that could confound associations. 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 the Netherlands that had moderate levels of starvation and similar experiences of war,

IN VITRO FERTILIZATION VARIATION

A recent novel design is attempting to disentangle prenatal from inherited effects by using pregnancies resulting from in vitro fertilization (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, it indicates that genetic confounding might be driving the association. The technique has been used to assess the impact of smoking during pregnancy. Rice and colleagues (2009) found that smoking was associated with reduced offspring birth weight regardless of whether the mother was biologically related to the offspring. A similar pattern of findings was also shown in a different sample (Thapar et al. 2009). However, the association between smoking and offspring antisocial behavior may have been dependent on inherited factors, as it was seen only in biologically related pairs. A later study that combined data using this study design with two studies that assessed adoption found converging evidence for an intrauterine effect of smoking on offspring conduct problems (Gaysina et al. 2013). These approaches provide evidence that may contribute to the triangulation of findings across a range of studies, but when considered in isolation do not allow for definitive decisions on causation.

but not quite the extreme caloric deprivation experienced in the most affected areas. Differential associations were still seen when using these cohorts as a control group, which suggests that other factors such as stress and the experience of wartime are unlikely to account for the results (Brown & 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 to allow for stronger causal inferences than are provided by conventional statistical adjustments. Negative control designs identify an exposure or outcome where no association is predicted but a similar confounding structure is shared 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 in which underlying confounding structures are likely to be very different, thus lessening 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 randomization, utilizes genetic variants as the proxy variables, which can also rule out reverse causation because genes are determined at conception. Family-based studies use shared genetic and environmental characteristics to generate highly matched case-control studies. Other methods can also be used in specific circumstances (for example, see the sidebar on In Vitro Fertilization Variation).

Although none of these techniques represents a panacea and each has 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 (2010b) combined cross-contextual and negative control designs 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 the United Kingdom, despite the different social patterns of smoking during pregnancy in 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.

The 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 the application of these methods to better understand the causal effects of intrauterine exposures to substances such as tobacco and alcohol 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 and offers great promise for future DOHaD research. A few key considerations can contribute to 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 hypothesized 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 those in cohort studies, and in Mendelian randomization studies, exposures are likely to be longer than in either RCT or cohort studies (as they may be present from soon after 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 DOHaD research. A risk factor may have an effect on an outcome only during a specific period of pregnancy, and if different studies measure variables at slightly different times, the timing rather than a lack of a causal association could be the reason for inconsistent results. Finally, multiple hypotheses may explain observed associations, and therefore applying principles of inference to the best explanation (Lipton 2004) and considering possible sources of bias will be important when attempting to triangulate results across different designs (Richmond et al. 2014).

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We 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. S.H.G. and M.R.M. are members of the UK Center for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Center of Excellence. We gratefully acknowledge funding from the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration.

LITERATURE CITED

Angrist JD, Pischke J-S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton Univ. Press

Attermann J, Obel C, Bilenberg N, Nordenbaek CM, 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 DJ. 1995. Fetal origins of coronary heart disease. *BMJ* 311:171–74

Barker DJ. 1997. Intrauterine programming of coronary heart disease and stroke. *Acta Paediatr.* 423:178–83

Barker DJ, Martyn CN. 1997. The fetal origins of hypertension. *Adv. Nephrol. Necker Hosp.* 26:65–72

Barker DJ, 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. *Int. J. Epidemiol.* 44:512–25

Brion MJ, Lawlor DA, Matijasevich A, Horta B, Anselmi L, et al. 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 MJ, Leary SD, Smith GD, Ness AR. 2007. Similar associations of parental prenatal smoking suggest child blood pressure is not influenced by intrauterine effects. *Hypertension* 49:1422–28

Brion MJ, Ness AR, Rogers I, Emmett P, Cribb V, et al. 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 MJ, Victora C, Matijasevich A, Horta B, Anselmi L, et al. 2010b. Maternal smoking and child psychological problems: disentangling causal and noncausal effects. *Pediatrics* 126:e57–65

Brown AS, Susser ES. 1997. Sex differences in prevalence of congenital neural defects after periconceptional famine exposure. *Epidemiology* 8:55–58

Brown AS, Susser ES. 2008. Prenatal nutritional deficiency and risk of adult schizophrenia. *Schizophr. Bull.* 34:1054–63

Burgess S, Timpson NJ, Ebrahim S, Davey Smith G. 2015. Mendelian randomization: Where are we now and where are we going? *Int. J. Epidemiol.* 44:379–88

Class QA, Rickert ME, Larsson H, Lichtenstein P, D'Onofrio BM. 2014. Fetal growth and psychiatric and socioeconomic problems: population-based sibling comparison. *Br. J. Psychiatry* 205:355–61

Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, et al. 2010. Illustrating bias due to conditioning on a collider. *Int. J. Epidemiol.* 39:417–20

D'Onofrio BM, Class QA, Lahey BB, Larsson H. 2014. Testing the developmental origins of health and disease hypothesis for psychopathology using family-based quasi-experimental designs. *Child Dev. Perspect.* 8:151–57

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 × environment interactions. *Perspect. Psychol. Sci.* 5:527–45

Davey Smith G. 2011. Epidemiology, epigenetics and the "gloomy prospect": embracing randomness in population health research and practice. *Int. J. Epidemiol.* 40:537–62

Davey Smith G. 2012a. Epigenesis for epidemiologists: Does evo-devo have implications for population health research and practice? *Int. J. Epidemiol.* 41:236–47

Davey Smith G. 2012b. Negative control exposures in epidemiologic studies. *Epidemiology* 23:350–51; author reply 351–52
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* 325:1437–38
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 AN1992. Confounding in epidemiological studies: why "independent" effects may not be all they seem. *BMJ* 305:757–59
Davey Smith G, Phillips AN, Neaton JD. 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, et al. 2015. Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study. *Lancet Psychiatry* 2:233–38
DiNardo JE, Pischke JS. 1997. The returns to computer use revisited: Have pencils changed the wage structure too? *Q. J. Econ.* 112:291–303
Ebrahim S, Davey Smith G. 2013. Should we always deliberately be non-representative? *Int. J. Epidemiol.* 42:1022–26
Enomoto N, Takase S, Yasuhara M, Takada A. 1991. Acetaldehyde metabolism in different aldehyde dehydrogenase-2 genotypes. *Alcohol. Clin. Exp. Res.* 15:141–44
Eriksson JG, 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.* 47:94–99
Eryigit Madzwamuse S, Baumann N, Jaekel J, Bartmann P, Wolke D. 2015. Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. *J. Child Psychol. Psychiatry* 56:857–64
Fall CH, Osmond C, Barker DJ, Clark PM, Hales CN, et al. 1995. Fetal and infant growth and cardiovascular risk factors in women. *BMJ* 310:428–32
Fewell Z, Davey Smith G, Sterne JA. 2007. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am. J. Epidemiol.* 166:646–55
Freathy RM, Ring SM, Shields B, Galobardes B, Knight B, et al. 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.* 18:2922–27
Gage SH, Munafò MR, MacLeod J, Hickman M, Davey Smith G. 2015. Cannabis and psychosis. *Lancet Psychiatry* 2:380
Gaysina D, Fergusson DM, Leve LD, Horwood J, Reiss D, et al. 2013. Maternal smoking during pregnancy and offspring conduct problems: evidence from 3 independent genetically sensitive research designs. *JAMA Psychiatry* 70:956–63
Hales CN, Barker DJ. 1992. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* 35:595–601
Hoek HW, Brown AS, Susser E. 1998. The Dutch famine and schizophrenia spectrum disorders. *Soc. Psychiatry Psychiatr. Epidemiol.* 33:373–79
Kelly YJ, Sacker A, Gray R, Kelly J, Wolke D, et al. 2012. Light drinking during pregnancy: still no increased risk for socioemotional difficulties or cognitive deficits at 5 years of age? *J. Epidemiol. Community Health* 66:41–48
Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, et al. 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: Eur. Monit. Cent. Drugs Drug Addict.
Kramer MS, Matush L, Vanilovich I, Platt RW, Bogdanovich N, et al. 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 *A Life Course Approach to Chronic Disease Epidemiology*, ed. D Kuh, Y Ben-Shlomo, pp. 15–37. Oxford, UK: Oxford Univ. Press. 2nd ed.

Lawlor DA, Davey Smith G, 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 DA, Timpson NJ, Harbord RM, Leary S, Ness A, et al. 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. *Sci. World J.* 2013:860915

Lipsitch M, Tchetgen Tchetgen E, Cohen T. 2010. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 21:383–88

Lipton P. 2004. *Inference to the Best Explanation.* London/New York: Routledge. 2nd ed.

Lorenz K. 1935. Der Kumpan in der Umwelt des Vogels. *J. Ornithol.* 83:137–213

Martin RM, Goodall SH, 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, et al. 2010. Potency trends of Δ9-THC and other cannabinoids in confiscated cannabis preparations from 1993 to 2008. *J. Forensic Sci.* 55:1209–17

Munafò MR, Timofeeva MN, Morris RW, Prieto-Merino D, Sattar N, et al. 2012. Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *J. Natl. Cancer Inst.* 104:740–48

Nilsson JP. 2014. *Alcohol availability, prenatal conditions, and long-term economic outcomes.* Work. Pap., Inst. Int. Econ. Stud., Stockholm Univ.

Oosterbeek H. 1997. Returns from computer use: a simple test on the productivity interpretation. *Econ. Lett.* 55:273–77

Patel R, Oken E, Bogdanovich N, Matush L, Sevkovskaya Z, et al. 2014. Cohort profile: the Promotion of Breastfeeding Intervention Trial (PROBIT). *Int. J. Epidemiol.* 43:679–90

Petitti DB, Perlman JA, Sidney S. 1986. Postmenopausal estrogen use and heart disease. *N. Engl. J. Med.* 315:131–32

Petitti DB, Perlman JA, Sidney S. 1987. Noncontraceptive estrogens and mortality: long-term follow-up of women in the Walnut Creek Study. *Obstet. Gynecol.* 70:289–93

Phillips AN, Davey Smith G. 1992. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat. Med.* 11:953–61

Pitkin RM. 2007. Folate and neural tube defects. *Am. J. Clin. Nutr.* 85:285–88S

Rice F, Harold GT, Boivin J, Hay DF, van den Bree M, Thapar A. 2009. Disentangling prenatal and inherited influences in humans with an experimental design. *PNAS* 106:2464–67

Richmond RC, Al-Amin A, Davey Smith G, Relton CL. 2014. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Hum. Dev.* 90:769–80

Roth C, Magnus P, Schjolberg S, Stoltenberg C, Suren P, et al. 2011. Folic acid supplements in pregnancy and severe language delay in children. *JAMA* 306:1566–73

Rothman K, Greenland S, Lash TL. 2008. *Modern Epidemiology*. Philadelphia, PA: 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, et al. 2013. Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. *JAMA* 309:570–77

Taylor AE, Davey Smith G, Bares CB, Edwards AC, Munafò MR. 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, et al. 2009. Prenatal smoking might not cause attention-deficit/hyperactivity disorder: evidence from a novel design. *Biol. Psychiatry* 66:722–27

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, et al. 2008. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452:638–42

Trezza V, Cuomo V, Vanderschuren LJ. 2008. Cannabis and the developing brain: insights from behavior. *Eur. J. Pharmacol.* 585:441–52

Tyrrell J, Huikari V, Christie JT, Cavadino A, Bakker R, et al. 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 RJ, 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 CG, Horta BL, Loret de Mola C, Quevedo L, Pinheiro RT, et al. 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 JJ, van den Bree MB, Munafò MR. 2011. Association of the *CHRNA5-A3-B4* gene cluster with heaviness of smoking: a meta-analysis. *Nicotine Tob. Res.* 13:1167–75

Ware JJ, van den Bree MB, Munafò MR. 2012. From men to mice: *CHRNA5/CHRNA3*, smoking behavior and disease. *Nicotine Tobacco Res.* 14:1291–99

Watson JB, Rayner R. 1920. Conditioned emotional reactions. *J. Exp. Psychol.* 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 SJ, Davey Smith G, Sayal K, Draper ES, et al. 2013. Prenatal alcohol exposure and offspring cognition and school performance. A "Mendelian randomization" natural experiment. *Int. J. Epidemiol.* 42:1358–70

Annual Review of Psychology

Volume 67, 2016

Contents