Fetal Origins of Mental Disorders? An Answer Based on Mendelian Randomization

Subhi Arafat and Camelia C. Minică
Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands

The Barker hypothesis states that low birth weight (BW) is associated with higher risk of adult onset diseases, including mental disorders like schizophrenia, major depressive disorder (MDD), and attention deficit hyperactivity disorder (ADHD). The main criticism of this hypothesis is that evidence for it comes from observational studies. Specifically, observational evidence does not suffice for inferring causality, because the associations might reflect the effects of confounders. Mendelian randomization (MR) — a novel method that tests causality on the basis of genetic data — creates the unprecedented opportunity to probe the causality in the association between BW and mental disorders in observation studies. We used MR and summary statistics from recent large genome-wide association studies to test whether the association between BW and MDD, schizophrenia and ADHD is causal. We employed the inverse variance weighted (IVW) method in conjunction with several other approaches that are robust to possible assumption violations. MR-Egger was used to rule out horizontal pleiotropy. IVW showed that the association between BW and MDD, schizophrenia and ADHD is not causal (all $p > .05$). The results of all the other MR methods were similar and highly consistent. MR-Egger provided no evidence for pleiotropic effects biasing the estimates of the effects of BW on MDD (intercept = -0.004, SE = 0.005, p = .372), schizophrenia (intercept = 0.003, SE = 0.01, p = .769), or ADHD (intercept = 0.009, SE = 0.01, p = .357). Based on the current evidence, we refute the Barker hypothesis concerning the fetal origins of adult mental disorders. The discrepancy between our results and the results from observational studies may be explained by the effects of confounders in the observational studies, or by the existence of a small causal effect not detected in our study due to weak instruments. Our power analyses suggested that the upper bound for a potential causal effect of BW on mental disorders would likely not exceed an odds ratio of 1.2.

■ **Keywords:** birth weight, depression, schizophrenia, attention deficit hyperactivity disorder, Mendelian randomization

The fetal origins hypothesis, also known as the Barker hypothesis, states that low birth weight (BW), an indication of poor fetal nutrition, is associated with increased risk of developing adult onset diseases. The British epidemiologist David Barker and his colleagues first suggested a potential role of fetal nutrition in the etiology of ischemic heart diseases after they observed that individuals in regions in England and Wales who showed increased fetal mortality between 1921 and 1925 also showed increased mortality rates from ischemic heart diseases several decades later (Barker & Osmond, 1986). In follow-up epidemiological studies using public records from Hertfordshire and Preston, Barker found that low BW was associated with mortality from ischemic heart disease, and also with type 2 diabetes and hypertension (Hales & Barker, 1992).

Attempts to explain the low BW diabetes association implicated insulin and other metabolic mediators in the causal mechanism. Subsequent studies investigating this relationship advanced the thrifty phenotype hypothesis, which proposed that poor fetal nutrition 'programs' the fetus, causing long-lasting effects on health. Accordingly, the prenatal period is critical for the subsequent development; poor nutrition during this period causes structural and functional adaptations, such as decreased insulin sensitivity, that enhance survival and prepare the fetus for postnatal life in a nutrient-poor environment. However, such changes would have negative effects in affluent environments, predisposing

to diabetes and metabolic syndrome (Hales & Barker, 1992, 2001).

Barker's hypothesis gained momentum in the following decades. Results were supported by several epidemiological and animal studies (Skogen & Øverland, 2012). The hypothesis also expanded to include other diseases, such as psychiatric disorders, and it was suggested that the role of poor prenatal environment might explain the associations between psychiatric and cardiometabolic diseases (Schlotz & Phillips, 2009).

Importantly, and in line with this hypothesis, the Dutch famine studies provided the earliest evidence for the role of prenatal nutrition in mental disorders. While an extreme case that reflects complex effects, those studies showed that individuals conceived during the famine showed increased risk of several psychiatric illnesses, particularly schizophrenia, and depression (Räikkönen et al., 2012; Schlotz & Phillips, 2009). Subsequent case-control and cohort studies, however, yielded mixed results. A recent systematic review (Wojcik et al., 2013) indicated that the association between low BW and depression is rather weak. For schizophrenia, studies to date have produced contradictory results (Abel et al., 2010; Gunnell et al., 2005; Schlotz & Phillips, 2009). The strongest evidence was provided for attention deficit hyperactivity disorder (ADHD), with several epidemiological studies consistently associating low BW with hyperactivity and inattention (O'Donnell & Meaney, 2016; Schlotz & Phillips, 2009).

Proponents of the Barker hypothesis suggest that the supporting evidence is strong, and that attention should shift toward revealing the causal mechanisms underlying the observed associations. However, the primary criticisms of the Barker hypothesis is that evidence for it comes from observational studies: observational evidence does not suffice for inferring causality, because the observed associations might reflect the effects of confounding variables (Skogen & Øverland, 2012). For instance, a meta-analysis of 55 studies investigating the relationship between BW and blood pressure suggested that the reported association could well be attributable to random error, reporting bias, or to other confounders (Huxley et al., 2002). Similarly, a systematic review, which established a weak association between BW and depression, also noted several limitations of the pool of studies, including publication bias and lack of adjustment for potential confounders (Wojcik et al., 2013).

Demonstrating that BW causally affects psychiatric traits, and estimating the size of such effects, would provide an impetus to health policies and would establish the prenatal period as a crucial therapeutic window. Alternatively, refuting the hypothesis that the association is causal will help us to avoid basing interventions on incorrect causality models. However, causal inference presents a challenge when tight experimental control is unfeasible or impossible. The question of whether BW causes adult psychiatric disorders cannot be addressed using a randomized controlled trial, as one cannot assign individuals randomly to different BW conditions to study its effects on psychiatric traits. Mendelian randomization (MR), a well-established method to demonstrate causality using genetic data, has created an unprecedented opportunity to probe the causality in the association between BW and mental disorders (Davey Smith & Ebrahim, 2003; Evans & Davey Smith, 2015; Pingault et al., 2018). That the genes assort randomly and independently (by Mendel's first and second law) serves a function that is similar to that of randomization in a randomized controlled trial. Specifically, one can form random groups of individuals based on their genes, given that the genes assort randomly and independently of other traits and of environmental factors that may typically confound observational studies (Burgess & Thompson, 2015; Evans & Davey Smith, 2015; Pingault et al., 2018). Therefore, by using genetic variants that are robustly associated with the risk factor/exposure instead of using the exposure itself, MR makes it possible to study causal relationships isolated from the effects of confounders. This allows us to draw conclusions about causality from observational studies. In short, MR is a very suitable technique to test for effects of BW.

Recent MR results have demonstrated a causal effect of BW on risk factors for coronary artery disease (Zanetti et al., 2018) and on type 2 diabetes (Wang et al., 2016). However, no study to date has employed this approach to study the causal effect of BW on psychiatric disorders. Psychiatric disorders are leading causes of disability worldwide, forming an enormous burden on the individual, family, and society. Reliably establishing a role of the prenatal environment in their pathogenesis may be relevant to public health policies (Skogen & Øverland, 2012). The present aim is to use MR to test whether BW has a causal effect on depression, schizophrenia, and ADHD. Our focus on these disorders is motivated by recent large genome-wide association studies (GWASs), which have established an increasing number of robust genetic associations. These provide reliable instruments that facilitate the application of the MR analysis.

Methods

We carried out two sample MR analyses to test the causal effects of BW on the risk of major depressive disorder (MDD), schizophrenia, and ADHD. We used publicly available summary statistics from GWAS conducted by the Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/results-and-downloads) and by the Early Growth Genetics Consortium (https://egg-consortium.org/).

MR allows one to probe causality in the relationships between risk factors/exposures and outcomes like mental health in non-experimental data (Davey Smith & Ebrahim, 2003). MR, similar to a randomized controlled trial, entails a 'randomization procedure' as it uses genetic variants (randomly inherited at conception) to group individuals with different levels of exposure. In fact, MR is a form of

instrumental variable (IV) analysis; IVs are variables directly associated with an exposure, with effects on the outcome assumed to be entirely mediated by the exposure (Lawlor et al., 2008). In MR, genetic variants (usually single nucleotide polymorphisms; SNPs) feature as IVs (Davey Smith & Ebrahim, 2003). To feature as a valid IV, a genetic variant should (a) be strongly associated with the exposure, (b) be uncorrelated with confounders (influences common to exposure and outcome), and (c) affect the outcome exclusively via the exposure (i.e., pleiotropic effects of the instrument on the outcome should be absent; Davey Smith & Ebrahim, 2003).

Power Analyses

We estimated statistical power for each MR analysis using the tool by Burgess (https://sb452.shinyapps.io/power/). In calculating the power, we used the following settings: a significance level of 0.05 and a coefficient of determination of 0.02 (in the regression of the exposure on the genetic variants used to instrument the analysis), and we considered two effect sizes — 20% and 10%, respectively, per standard deviation change in the exposure (i.e., $OR = 1.2$ and $OR = 1.1$).

Mendelian Randomization

The inverse variance weighted (IVW) procedure was employed to probe causality (Burgess et al., 2013). The IVW causal effect estimate is calculated based on multiple SNPs and on the ratio of coefficients method. We used genetic variants from the largest GWASs of BW (Horikoshi et al., 2016), depression (Wray et al., 2018), schizophrenia (Ripke et al., 2013), and ADHD (Demontis et al., 2017). The 60 SNPs associated with BW at a genome-wide significance level in the offspring genotype analysis ($p < 5 \times 10^{-8}$) were used as instruments. For each genetic variant i ($i = 1...N_{SNP}$), we computed the Wald ratio estimate as (Burgess et al., 2013)

$$\beta_{IVi} = \beta o_{SNPi}/\beta_{ESNPi} \qquad (1)$$

where βo_{SNPi} is the regression coefficient in the outcome on SNPi regression, and β_{ESNPi} is the regression coefficient in the exposure on SNPi regression. As one does in a meta-analysis, we combined these ratios of coefficients by weighting them by their inverse variance (Burgess et al., 2013). Effectively, using the IVW procedure, one regresses the vector of NSNP associations with the outcome on the vector of associations with the exposure, while fixing the intercept to zero and employing inverse variance weighting. The IVW produces unbiased estimates of causal effects as long as the SNPs employed are valid IVs (Bowden et al., 2015). Assumptions (b) and (c) stated above are not empirically testable, and are unlikely to hold when many variants are used as instruments (as this increases the probability of pleiotropy; Bowden et al., 2016). To avoid the bias due to potential violation of the 'no horizontal pleiotropy' assump-

tion, we used the following MR methods, which are known to be robust in the presence of invalid IVs: median- (Bowden et al., 2016) and mode-based methods (Hartwig et al., 2017), and the MR-Egger regression (Bowden et al., 2015). Furthermore, we used forest plots (Wickham, 2010) to visualize the causal estimates based on each individual instrument, and the combined causal estimates (Hartwig et al., 2017).

The simple median estimator is calculated as the median of the set of ratio coefficients from each SNP selected to instrument the analysis. Even if up to half of these SNPs are invalid instruments (i.e., pleiotropic), the simple median method will produce unbiased estimates of the causal effect (Bowden et al., 2016). Unlike the simple median — where the ratio estimates from all instruments receive equal weights — the weighted median estimator weights the coefficients by their inverse variance to place more importance on instruments with more precise estimates. The resulting causal estimate is valid as long as at least half the weights are based on valid instruments (Bowden et al., 2016). The mode-based estimator is calculated as the most frequent estimate (i.e., the mode) of the set of ratio coefficients estimated from each genetic instrument. In this approach, the strong 'no pleiotropy' assumption is replaced with the assumption that the largest group of SNPs yielding similar estimates of the causal effect includes solely non-pleiotropic instruments (Hartwig et al., 2017). The weighted mode is similar to the simple mode, except that individual ratio estimates are weighted (Hartwig et al., 2017).

We also used MR-Egger to correct for potential horizontal pleiotropic effects (Bowden et al., 2015). MR-Egger, similar to the IVW method, performs a weighted linear regression using inverse variance weights; yet unlike the IVW method, it freely estimates the intercept to capture potential pleiotropic effects. As such, under the weaker assumption that the effects of the SNPs on the exposure are uncorrelated with the effects of the SNPs on the outcome, MR-Egger is expected to provide unbiased causal estimates even when all of the SNPs display pleiotropy (Bowden et al., 2015; see also Burgess & Thompson, 2017, for more details on this procedure). Finally, we used funnel plots as a visual test for horizontal pleiotropy, where symmetry is indicative of lower probability of pleiotropy (Bowden et al., 2015).

Tests of Heterogeneity

The causal estimates yielded by multiple instruments are expected to vary only by chance, if the SNPs satisfy the IV assumptions and the SNPs have the same causal effect size. Large inter-instrument heterogeneity may be indicative of pleiotropic effects (Bowden et al., 2017). To assess heterogeneity, we used Cochran's Q test. Cochran's Q test is commonly employed in meta-analyses to test whether the observed discrepancy between individual estimates is consistent with sampling variation; the test is also useful to assess heterogeneity in the IVW model (Greco et al., 2015).

TABLE 1

Description of the samples used in the genome-wide association studies of the exposure (birth weight) and of the outcomes considered in the Mendelian randomization analyses, and the statistical power to detect effect sizes of 20% ($OR = 1.2$) and 10% ($OR = 1.1$) per one standard deviation change in birth weight, given an alpha of 0.05

Phenotype	Sample size	Cases/controls	Power (effect size $OR = 1.2$)	Power (effect size $OR = 1.1$)	Publication
BW	153,781	–	–	–	Horikoshi et al. (2016)
MDD	173,005	59,851/113,154	99%	76%	Wray et al. (2018)
Schizophrenia	79,845	34,241/45,604	95%	47%	Ripke et al. (2013)
ADHD	55,374	20,183/35,191	83%	33%	Demontis et al. (2017)

Note: BW = birth weight; MDD = major depressive disorder; ADHD = attention deficit hyperactivity disorder.
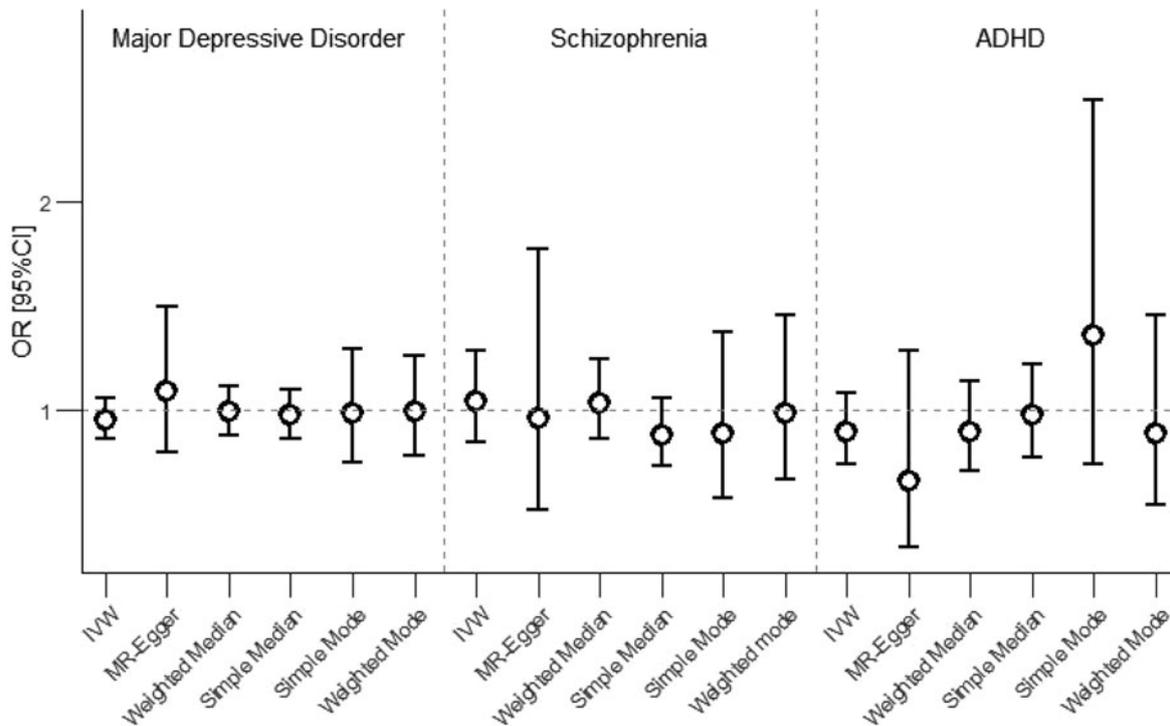
FIGURE 1

Results of the Mendelian randomization analyses testing causality in the association between birth weight and major depressive disorder, schizophrenia, and attention deficit hyperactivity disorder (ADHD). Error bars represent the 95% confidence intervals. Note: IVW = inverse variance weighted.

Forest plots were also used to visually examine the degree of heterogeneity in causal estimates based on each individual instrument (where non-overlapping confidence intervals indicate heterogenous effects).

Results

Power Analyses

Statistical power in a binary outcome MR analysis depends on sample size, cases-to-controls ratio, and on the coefficient of determination of exposure on the instruments (R^2; Bowden et al., 2015). The SNPs used to instrument the MR analysis explained ~2% of the variance in BW (Horikoshi et al., 2016). Details on power, sample sizes and data sources are provided in Table 1.

The power analyses showed that our analyses were adequately powered (>80%) to detect a causal effect $OR = 1.2$

given an alpha of 0.05. The power dropped under this optimal level when considering smaller effect sizes ($OR = 1.1$).

Mendelian Randomization Analyses

Figure 1 displays the MR results.

All the MR methods showed no evidence for a causal effect of BW on any of the outcomes (all 95% confidence intervals include odds ratio of 1). Figure S1 shows the MR estimates based on the individual instruments, and the combined causal estimates produced using the different two-sample MR estimators.

Tests of Horizontal Pleiotropy: MR-Egger and Funnel Plots

The results of MR-Egger regression, which corrects for horizontally pleiotropic effects, also showed no evidence for a
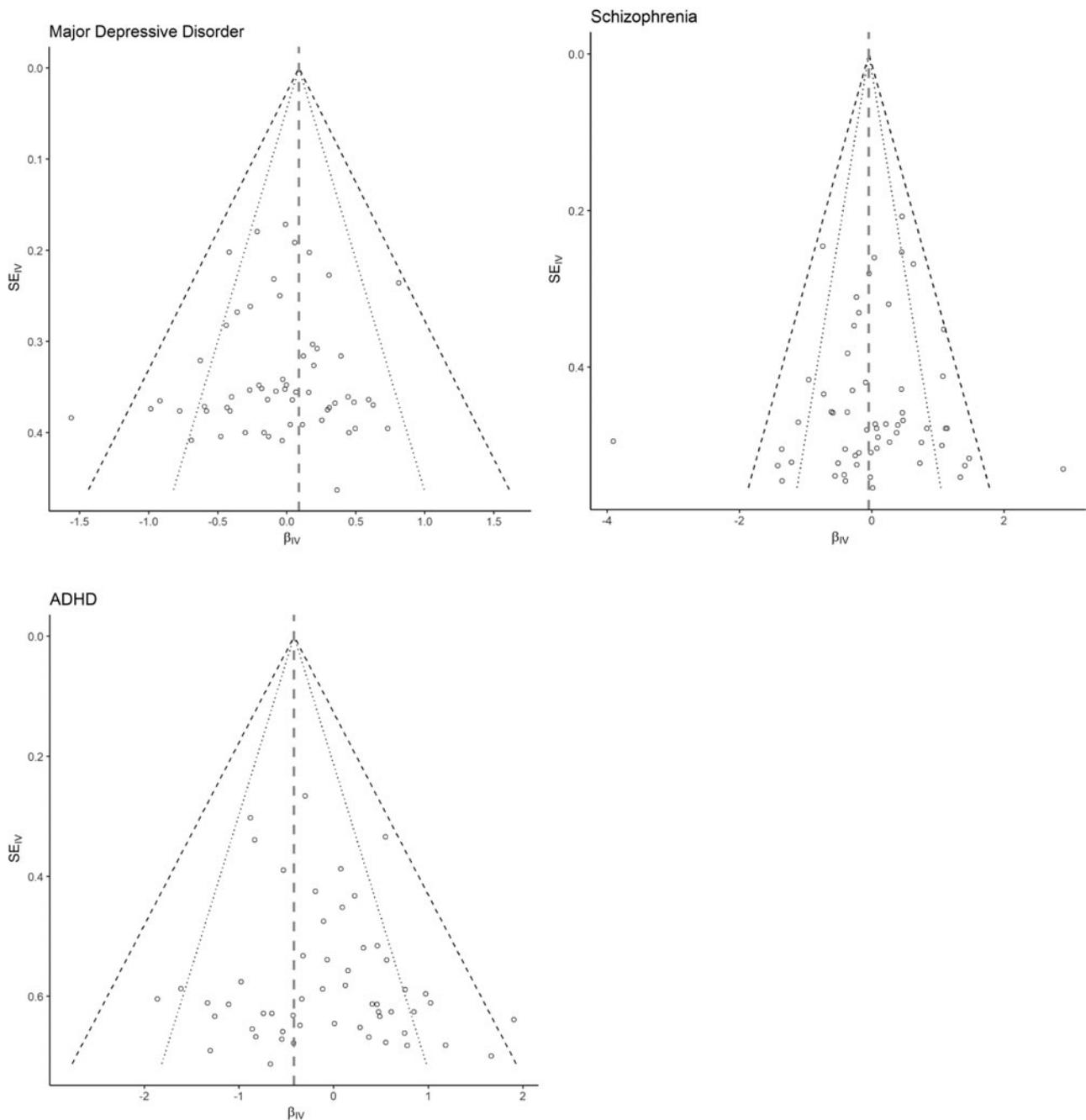
FIGURE 2

Forrest plots showing the relationship between causal estimates using each individual instrument (log odds) and the standard error of the estimate. The effects estimated based on MR-Egger are displayed as dashed grey vertical lines. The black dotted lines correspond to the lower and upper limits of the 95% confidence interval region, the black dashed lines corresponding to the 99% confidence interval region. Note: ADHD = attention deficit hyperactivity disorder.

causal effect of BW on any of the psychiatric disorders. Additionally, we could not reject the null hypothesis of no horizontal pleiotropy, as the test of the MR-Egger intercept was not significant; this result suggests that the assumption of 'no horizontal pleiotropy' holds. This is also demonstrated by funnel plots shown in Figure 2.

The plots were mostly symmetrical, which is consistent with the results of MR-Egger pleiotropy test.

Tests of Heterogeneity
The Cochran's Q indicated that the estimates based on the individual instruments are heterogenous (MDD: Cochran's

$Q(57) = 97.554$, $p = 7e\text{-}04$; Schizophrenia: Cochran's $Q(57)$ $= 231.647$, $p = 6.46e\text{-}23$; ADHD: Cochran's $Q(52) = 97.72$, $p = .0001$). The large inter-instrument variation is also evident in the forest plots in Figure S1.

Discussion

In this study, we used the two-sample MR procedure to test the Barker hypothesis concerning the fetal origins of adult mental disorders. We assessed whether the observed epidemiological associations between BW, on the one hand, and ADHD, MDD, and schizophrenia, on the other, are causal. We used several methods that are robust to a certain degree to violation of the 'no pleiotropy' assumption, hence providing a good means to check the validity of our results. Furthermore, we used several diagnostic tests that can detect bias resulting from potential assumption violation.

Our findings do not support a causal effect of BW on any of the outcomes that we considered. The results were consistent across the methods. However, the interpretation of these results hinges upon the tenability of all MR assumptions. The first assumption (that the instrument associates robustly with the exposure variable) holds true, as the SNPs that we used as IV explained a significant proportion of the variance in BW (around 2%; Horikoshi et al., 2016). The second assumption (that the instrument is independent of confounders) cannot be tested rigorously because of the many potential confounders. According to Schlotz and Phillips (2009), socio-economic status, education, and maternal smoking may lead to an association between BW and mental disorders. To get an indication of whether the SNPs employed to instrument the current analyses associate with these potential confounders, we used summary statistics obtained from the GWASs of educational attainment (Okbay et al., 2016) and smoking behavior (Furberg et al., 2010). There were 26 and 59 BW-associated variants that passed the quality control checks in the GWAS of smoking and in the GWAS of educational attainment, respectively. Of these, we identified only one SNP that passed the significance threshold of 0.05 in the GWAS of smoking, and 12 SNPs that associated with educational attainment (see Table S1 for details). We note that this alpha threshold is overly liberal given the large number of multiple comparisons. We wanted to maximize the power to identify even very small genetic associations with the potential confounders. To check the effects of those correlations on our results, we re-ran the MR analyses after removing any SNP that showed a significant association with any of the confounders (13 SNPs removed). Excluding these SNPs from the analyses did not change the results and the current conclusions (see Figure S2). Davey-Smith and colleagues (Smith, 2011; Smith et al., 2007) also provided empirical support for the assumption that the genetic variants are distributed in the population independent of behavioral,

social, and physiological factors that might confound epidemiological studies.

The MR-Egger pleiotropy test found no evidence for pleiotropic effects, although our results showed a high degree of heterogeneity (which might be indicative of horizontal pleiotropy). To further test this assumption, we used as an alternative test the recently developed MR-PRESSO (MR pleiotropy residual sum and outlier; Verbanck et al., 2018). The MR-PRESSO conducts a global test to detect overall pleiotropy; next, variants yielding outlying causal effects are removed, as such SNPs likely have pleiotropic effects. While the test demonstrated that such outliers are present, correcting for them still produced no evidence for a causal effect (see Table S2).

Another important consideration when interpreting the results is statistical power. Our power analyses indicated that we had relatively good power to detect an effect as small as an odds ratio of 1.1 in the BW–MDD study, probably owing to the large sample used in the GWAS of MDD ($N = 173{,}005$ individuals). The power was lower in the other analyses. The MR studies testing the causal association between BW and ADHD and schizophrenia had adequate power to detect effects as large as an odds ratio of 1.2 but not an effect of an odds ratio of 1.1.

To put these results into perspective, we note that Thompson et al. (2001) demonstrated an inverse association between low BW and men's adult depression (Thompson et al., 2001), while Gale and Martyn (2004) showed that this association is observed in both men and women with very low BW (<2.5 kg). Similarly, in a recent meta-analysis of 14 observational studies, De Mola et al. (2014) found an association between low BW (<2.5 kg) and depression in adults ($OR = 1.39$, 95% CI [1.21, 1.60]). They further showed that the strength of the effect varied over the studies as a function of (a) the threshold used to define BW categories, (b) gender composition of the sample, (c) presence or absence of adjustment for potential confounders (e.g., socio-economic status, gestational age), (d) type of study design, and (e) age of the participants. Conversely, a meta-analysis of 18 studies by Wojcik et al. (2013) found a weak effect ($OR = 1.15$, 95% CI [1.00, 1.32]) of low BW on depression or psychological distress (they obtained a similar effect when restricting the analysis to the 15 studies that reported only depression as the outcome). Yet, correction for publication bias rendered the weak association observed by Wojcik et al. no longer significant. Similar null associations were observed in samples either restricted to women (Inskip et al., 2008) or to men (Osler et al., 2005). The results concerning the relationship between BW and schizophrenia are also inconsistent. Several studies supported an association in individuals weighing less than 2.5 kg at birth (Cannon et al., 2002; Gunnell et al., 2003), while subsequent well-powered observational studies conducted in the large Scandinavian databases showed that the negative association remains significant when considering the normal BW

range (up to 4.5 kg or more; Abel et al., 2010; Eide et al., 2013). There is also evidence of no association between BW and schizophrenia (see, e.g., Gunnell et al., 2005), as well as evidence supporting a reverse J-shaped association, with the largest odds of developing schizophrenia observed in individuals with very small weight at birth (<2.5 kg; see, e.g., Gunnell et al., 2003; Moilanen et al., 2010). Unlike the present study, studies to date investigating the relationship between BW and mental disorders like schizophrenia and MDD were based on observational evidence and therefore cannot infer causality. It is likely that the previously reported associations reflect the effects of confounders, as adjustment for known confounders does not suffice (Pingault et al., 2018); specifically, interpreting these associations as causal requires exhaustive adjustment for all possible confounders, as well as the untestable assumption of absence of reverse causation. Our study, as the first to use the MR approach, which has distinct advantages such as increased ecologic validity and absence of confounding, does not support a causal effect of BW on mental disorders. Our results, however, do not exclude the possibility of a causal association between (extremely) low BW (Lærum et al., 2017) and schizophrenia and more severe forms of depression (i.e., recurrent severe depressive symptoms; Colman et al., 2007; Wojcik et al., 2013). Alternatively, it is possible that there is a small effect that went undetected in our study due to insufficient statistical power to detect relatively weak causal associations.

On the other hand, the evidence for the association between low BW and ADHD produced by studies to date is more robust (Momany et al., 2018; Wiles et al., 2006). Importantly, there are also several studies that probed the causality in this association using the co-twin control method; these studies consistently demonstrated an effect of BW on ADHD symptoms and attention problems (Ficks et al., 2013; Groen-Blokhuis et al., 2011; Pettersson et al., 2015). The co-twin method probes the causal effect of a risk factor on an outcome or disorder by comparing the within-pair mean differences/relative risk for developing the disorder between unrelated participants, and monozygotic and dizygotic twin pairs discordant for exposure to the risk factor (Hart et al., 2013; Kendler et al., 1993; Middeldorp et al., 2008). While these studies employed a continuous measure of attention problems, our study was based on summary statistics obtained in the GWAS of ADHD (Demontis et al., 2017), which employed a dichotomous outcome (the presence or absence of ADHD, assessed based on a diagnostic interview or reported by parents or teachers). Using a continuous outcome confers larger statistical power relative to using a dichotomous measure (Groen-Blokhuis et al., 2011). Again, our results do not exclude the possibility that there is a small causal effect that was not detected in our study due to weak instruments and the use of a dichotomous phenotype. Another possible explanation of the differences in the results is that the co-twin control method

makes strong assumptions concerning the environmental influences not shared within a twin pair (Hart et al., 2013). It is plausible that there are intrauterine factors not controlled for by this design — such as the difference in blood flow between twins — that could confound the relationship between BW and attention problems.

The Barker hypothesis suggests that poor fetal environment is linked to increased risk of adult diseases. BW is often used as a proxy for fetal development; yet, it might be possible that BW is a poor indicator of intrauterine factors contributing to the later development of mental disorders. To explore this possibility, we tested for causal effects using another indicator of fetal development — gestational age — by employing the same MR methods. These analyses also produced no evidence for a causal effect of gestational age on mental disorders (see Figure S3). These findings do provide additional support for our results; however, it must be noted that these new analyses had limitations, such as the small number of available instrumental SNPs, low proportion of variance explained in the exposure (weak instruments), and low statistical power.

Strengths and Limitations

To our knowledge, this is the first study to evaluate the effect of BW on mental disorders by using new statistical methods that can test causal hypothesis on the basis of genetic data. MR methods provide several advantages over the observational analysis, such as producing a causal effect estimate free from the effects of confounders (Lawlor et al., 2008; Pingault et al., 2018). Furthermore, the study applied various methods, each employing different assumptions. The fact that these different approaches yielded consistent results increases the likelihood that our results are robust. In addition, this study is the first to make use of summary statistics of the recently published GWASs to test the fetal origins of mental disorders.

There are three potential limitations that require attention in the interpretation of the current results. Note that the SNPs we used as IV explained only 2% of the variance in BW, which means that the instruments were relatively weak. Although we used multiple SNPs to test the causal hypotheses, these variants were employed individually; hence, the approach may still be vulnerable to weak instrument bias relative to an approach that uses a polygenic score to instrument the analysis. It is known that in the two-sample analyses, weak instruments bias the estimate toward the null (Evans & Davey Smith, 2015). Hence, it is worth readdressing the Barker hypothesis as novel GWAS summary statistics become available (particularly the BW-ADHD and BW-schizophrenia relationships, as the current study had low power to detect small effects), and by employing alternative approaches that allow for the use of strong instruments in the form of polygenic scores (Minică et al., 2018). Another limitation concerns the relatively low power of MR-Egger to identify pleiotropic effects (Verbanck et al., 2018).

Although MR-PRESSO is more powerful, its power depends heavily on the proportion of pleiotropic variants that made up the instrument. MR-PRESSO has good power if at least 10% of the variants have pleiotropic effects (Verbanck et al., 2018). However, a single pleiotropic variant is sufficient to bias the MR results. One final issue is that the assessment of outcomes was not always based on a clinical diagnosis; this might potentially affect the accuracy of the results.

Conclusion

Based on the current findings, we found no support for the Barker hypothesis concerning the fetal origins of mental disorders. To account for lack of power to identify small effects, it is important to re-run the analysis when more SNPs associated with BW are identified, resulting in stronger instruments. One way to further explore pleiotropy is to combine MR with twin models using the MR-direction of causation model (Minică et al., 2018). In addition, indicators for fetal development other than BW can be used as proxies, such as height at birth and head circumference.

Acknowledgments

Conflict of Interest

The authors have no conflict of interest to declare.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/thg.2018.65

References

Abel, K. M., Wicks, S., Susser, E. S., Dalman, C., Pedersen, M. G., Mortensen, P. B., & Webb, R. T. (2010). Birth weight, schizophrenia, and adult mental disorder: Is risk confined to the smallest babies? *Archives of General Psychiatry, 67*, 923–930.

Barker, D. J., & Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *The Lancet, 327*, 1077–1081.

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, 44*, 512–525.

Bowden, J., Davey Smith, G., Haycock, P. C., & Burgess, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology, 40*, 304–314.

Bowden, J., Del Greco, M. F., Minelli, C., Davey Smith, G., Sheehan, N., & Thompson, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine, 36*, 1783–1802.

Burgess, S., Butterworth, A., & Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology, 37*, 658–665.

Burgess, S., & Thompson, S. G. (2015). *Mendelian randomization: Methods for using genetic variants in causal estimatio*. London, UK: Chapman & Hall/CRC Press.

Burgess, S., & Thompson, S. G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology, 32*, 377–389.

Cannon, M., Jones, P. B., & Murray, R. M. (2002). Obstetric complications and schizophrenia: Historical and meta-analytic review. *American Journal of Psychiatry, 159*, 1080–1092.

Colman, I., Ploubidis, G. B., Wadsworth, M. E., Jones, P. B., & Croudace, T. J. (2007). A longitudinal typology of symptoms of depression and anxiety over the life course. *Biological Psychiatry, 62*, 1265–1271.

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?. *International Journal of Epidemiology, 32*, 1–22.

De Mola, C. L., De França, G. V. A., de Avila Quevedo, L., & Horta, B. L. (2014). Low birth weight, preterm birth and small for gestational age association with adult depression: Systematic review and meta-analysis. *The British Journal of Psychiatry, 205*, 340–347.

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J., Grove, J., Hansen, C. S., Hauberg, M. E., Hollegaard, M. V., Howrigan, D. P., Huang, H., Maller, J., Martin, A. R., Moran, J., Pallesen, J., Palmer, D. S., Pedersen, C. B., Pedersen, M. G., Poterba, T., Poulsen, J. B., Ripke, S., Robinson, E. B., Satterstrom, K. F., Stevens, C., Turley, P., Won, H., ADHD Working Group of the Psychiatric Genomics Consortium (PGC), Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team, Andreassen, O. A., Burton, C., Boomsma, D., Cormand, B., Dalsgaard, S., Franke, B., Gelernter, J., Geschwind, D., Hakonarson, H., Haavik, J., Kranzler, H., Kuntsi, J., Langley, K., Lesch, K.-P., Middeldorp, C., Reif, A., Rohde, L. A., Roussos, P., Schachar, R., Sklar, P., SonugaBarke, E., Sullivan, P. F., Thapar, A., Tung, J., Waldman, I., Nordentoft, M., Hougaard, D. M., Werge, T., Mors, O., Mortensen, P. B., Daly, M. J., Faraone, S. V., Børglum, A. D., Neale, B. M. & Cerrato, F. (2017). Discovery of the first genome-wide significant risk loci for ADHD. *BioRxiv*, 145581.

Eide, M., Moster, D., Irgens, L., Reichborn-Kjennerud, T., Stoltenberg, C., Skjaerven, R., … Abel, K. (2013). Degree of fetal growth restriction associated with schizophrenia risk in a national cohort. *Psychological Medicine, 43*, 2057–2066.

Evans, D. M., & Davey Smith, G. (2015). Mendelian randomization: New applications in the coming age of hypothesis-free causality. *Annual Review of Genomics and Human Genetics, 16*, 327–350.

Ficks, C. A., Lahey, B. B., & Waldman, I. D. (2013). Does low birth weight share common genetic or environmental risk with childhood disruptive disorders?. *Journal of Abnormal Psychology, 122*, 842–853.

Furberg, H., Kim, Y., Dackor, J., Boerwinkle, E., Franceschini, N., Ardissino, D., … Merlini, P. A. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics, 42*, 441–447.

Gale, C. R., & Martyn, C. N. (2004). Birth weight and later risk of depression in a national birth cohort. *The British Journal of Psychiatry, 184*, 28–33.

Greco, M. F. D., Minelli, C., Sheehan, N. A., & Thompson, J. R. (2015). Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine, 34*, 2926–2940.

Groen-Blokhuis, M. M., Middeldorp, C. M., van Beijsterveldt, C. E., & Boomsma, D. I. (2011). Evidence for a causal association of low birth weight and attention problems. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*, 1247–1254.

Gunnell, D., Harrison, G., Whitley, E., Lewis, G., Tynelius, P., & Rasmussen, F. (2005). The association of fetal and childhood growth with risk of schizophrenia. Cohort study of 720,000 Swedish men and women. *Schizophrenia Research, 79*, 315–322.

Gunnell, D., Rasmussen, F., Fouskakis, D., Tynelius, P., & Harrison, G. (2003). Patterns of fetal and childhood growth and the development of psychosis in young males: A cohort study. *American Journal of Epidemiology, 158*, 291–300.

Hales, C. N., & Barker, D. J. (1992). Type 2 (non-insulin-dependent) diabetes mellitus: The thrifty phenotype hypothesis. *Diabetologia, 35*, 595–601.

Hales, C. N., & Barker, D. J. (2001). The thrifty phenotype hypothesis. *British Medical Bulletin, 60*, 5–20.

Hart, S. A., Taylor, J., & Schatschneider, C. (2013). There is a world outside of experimental designs: Using twins to investigate causation. *Assessment for Effective Intervention, 38*, 117–126.

Hartwig, F. P., Davey Smith, G., & Bowden, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology, 46*, 1985–1998.

Horikoshi, M., Beaumont, R. N., Day, F. R., Warrington, N. M., Kooijman, M. N., Fernandez-Tajes, J.,… Grarup, N. (2016). Genome-wide associations for birth weight and correlations with adult disease. *Nature, 538*, 248–252.

Huxley, R., Neil, A., & Collins, R. (2002). Unravelling the fetal origins hypothesis: Is there really an inverse association between birthweight and subsequent blood pressure?. *The Lancet, 360*, 659–665.

Inskip, H. M., Dunn, N., Godfrey, K. M., Cooper, C., Kendrick, T., & Southampton Women's Survey Study, G. (2008). Is birth weight associated with risk of depressive symptoms in young women? Evidence from the Southampton women's survey. *American Journal of Epidemiology, 167*, 164–168.

Kendler, K. S., Neale, M. C., MacLean, C. J., Heath, A. C., Eaves, L. J., & Kessler, R. C. (1993). Smoking and major depression: A causal analysis. *Archives of General Psychiatry, 50*, 36–43.

Lærum, A. M., Reitan, S. K., Evensen, K. A. I., Lydersen, S., Brubakk, A.-M., Skranes, J., & Indredavik, M. S. (2017). Psychiatric disorders and general functioning in low birth weight adults: A longitudinal study. *Pediatrics*, *27*(8): 1133–63. e20162135.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine, 27*, 1133–1163.

Middeldorp, C., Cath, D., Beem, A., Willemsen, G., & Boomsma, D. (2008). Life events, anxious depression and personality: A prospective and genetic study. *Psychological Medicine, 38*, 1557–1565.

Minică, C. C., Dolan, C. V., Boomsma, D. I. *et al.* (2018). *Behav Genet 48*(8): 337. https://doi.org/10.1007/s10519-018-9904-4.

Minică, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E., & Neale, M. C. (2018). Extending causality tests with genetic instruments: An integration of Mendelian randomization with the classical twin design. *Behavior Genetics*, 1–13.

Moilanen, K., Jokelainen, J., Jones, P. B., Hartikainen, A.-L., Järvelin, M.-R., & Isohanni, M. (2010). Deviant intrauterine growth and risk of schizophrenia: A 34-year follow-up of the Northern Finland 1966 Birth Cohort. *Schizophrenia Research, 124*, 223–230.

Momany, A. M., Kamradt, J. M., & Nikolas, M. A. (2018). A meta-analysis of the association between birth weight and attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology, 46*, 1409–1426.

O'Donnell, K. J., & Meaney, M. J. (2016). Fetal origins of mental health: The developmental origins of health and disease hypothesis. *American Journal of Psychiatry, 174*, 319–328.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … Meddens, S. F. W. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*, 539–542.

Osler, M., Nordentoft, M., & Andersen, A.-M. N. (2005). Birth dimensions and risk of depression in adulthood: Cohort study of Danish men born in 1953. *The British Journal of Psychiatry, 186*, 400–403.

Pettersson, E., Sjölander, A., Almqvist, C., Anckarsäter, H., D'Onofrio, B. M., Lichtenstein, P., & Larsson, H. (2015). Birth weight as an independent predictor of ADHD symptoms: A within-twin pair analysis. *Journal of Child Psychology and Psychiatry, 56*, 453–459.

Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, *19*, 566–580.

Räikkönen, K., Pesonen, A.-K., Roseboom, T. J., & Eriksson, J. G. (2012). Early determinants of mental health. *Best Practice & Research Clinical Endocrinology & Metabolism, 26*, 599–611.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., … Fromer, M. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics, 45*, 1150–1159.

Schlotz, W., & Phillips, D. I. (2009). Fetal origins of mental health: Evidence and mechanisms. *Brain, Behavior, and Immunity, 23*, 905–916.

Skogen, J. C., & Øverland, S. (2012). The fetal origins of adult disease: A narrative review of the epidemiological literature. *JRSM Short Reports, 3*, 1–7.

Smith, G. D. (2011). Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. *Genes & Nutrition, 6*, 27–43.

Smith, G. D., Lawlor, D. A., Harbord, R., Timpson, N., Day, I., & Ebrahim, S. (2007). Clustered environments and randomized genes: A fundamental distinction between conventional and genetic epidemiology. *PLoS Medicine, 4*, e352.

Thompson, C., Syddall, H., Rodin, I., Osmond, C., & Barker, D. J. (2001). Birth weight and the risk of depressive disorder in late life. *The British Journal of Psychiatry, 179*, 450–455.

Verbanck, M., Chen, C.-Y., Neale, B., & Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics, 50*, 693.

Wang, T., Huang, T., Li, Y., Zheng, Y., Manson, J. E., Hu, F. B., & Qi, L. (2016). Low birthweight and risk of type 2 diabetes: A Mendelian randomisation study. *Diabetologia, 59*, 1920–1927.

Wickham, H. (2010). ggplot2: Elegant graphics for data analysis. *Journal of Statistical Software, 35*, 65–88.

Wiles, N. J., Peters, T. J., Heron, J., Gunnell, D., Emond, A., & Lewis, G. (2006). Fetal growth and childhood behavioral problems: Results from the ALSPAC cohort. *American Journal of Epidemiology, 163*, 829–837.

Wojcik, W., Lee, W., Colman, I., Hardy, R., & Hotopf, M. (2013). Foetal origins of depression? A systematic review and meta-analysis of low birth weight and later depression. *Psychological Medicine, 43*, 1–12.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., … Andlauer, T. M. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics, 50*, 668–681.

Zanetti, D., Tikkanen, E., Gustafsson, S., Priest, J. R., Burgess, S., & Ingelsson, E. (2018). Birthweight, type 2 diabetes mellitus, and cardiovascular disease: Addressing the Barker hypothesis with Mendelian randomization. *Circulation: Genomic and Precision Medicine, 11*, e002054.