



Bounding the average causal effect in Mendelian randomisation studies with multiple proposed instruments: An application to prenatal alcohol exposure and attention deficit hyperactivity disorder

Elizabeth W. Diemer^{1,2,3} | Alexandra Havdahl^{4,5,6} | Ole A. Andreassen⁷ |
 Marcus R. Munafò^{4,8,9} | Pal R. Njolstad^{10,11} | Henning Tiemeier^{1,12} | Luisa Zuccolo^{4,13,14} |
 Sonja A. Swanson^{2,3,15,16}

¹Department of Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, the Netherlands

²CAUSALab, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

³Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

⁴MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

⁵Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway

⁶Department of Mental Disorders, Norwegian Institute of Public Health

⁷NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway

⁸School of Psychological Science, University of Bristol, Bristol, UK

⁹NIHR Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, UK

¹⁰Department of Paediatric and Adolescent Medicine, Haukeland University Hospital, Bergen, Norway

¹¹KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway

¹²Department of Social and Behavioral Science, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

¹³Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

¹⁴Centre for Health Data Science, Human Technopole Foundation, Milan, Italy

¹⁵Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands

¹⁶Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence

Elizabeth W. Diemer, Department of Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, the Netherlands.

Email: eld883@mail.harvard.edu

Funding information

Department of Veterans Affairs Cooperative Studies Program; National Institute of Alcohol Abuse and Alcoholism; Norwegian Institute of Education and Research; Research Council of Norway (NRC), Grant/Award Number: 274611, 223273 and 229624; Stiftelsen Kristian Gerhard Jebsen; Trond Mohn Foundation; Novo Nordisk Foundation; University

Abstract

Background: As large-scale observational data become more available, caution regarding causal assumptions remains critically important. This may be especially true for Mendelian randomisation (MR), an increasingly popular approach. Point estimation in MR usually requires strong, often implausible homogeneity assumptions beyond the core instrumental conditions. Bounding, which does not require homogeneity assumptions, is infrequently applied in MR.

Objectives: We aimed to demonstrate computing nonparametric bounds for the causal risk difference derived from multiple proposed instruments in an MR study where effect heterogeneity is expected.

Social Media Quote In large Mendelian randomisation studies, nonparametric bounds may show us how much we do not know.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Paediatric and Perinatal Epidemiology* published by John Wiley & Sons Ltd.

of Bergen; Western Norway health Authorities (Helse Vest); Norwegian Ministry of Health and Care Services; Ministry of Education and Research; South-Eastern Norway Regional Health Authority, Grant/Award Number: 2018059; Marie Skłodowska-Curie, Grant/Award Number: 721567; NWO/ZonMW Veni, Grant/Award Number: 91617066; Dutch Ministry of Education, Culture, and Science; Netherlands Organisation for Scientific Research, Grant/Award Number: 024.001.003; MRC Integrative Epidemiology Unit at the University of Bristol, Grant/Award Number: MC_UU_00011/1 and MC_UU_00011/7; UK Medical Research Council; Wellcome, Grant/Award Number: 217065/Z/19/Z; University of Bristol; MRC, Grant/Award Number: G0902144; 23andMe; NIHR Bristol Biomedical Research Centre

Methods: Using data from the Norwegian Mother, Father and Child Cohort Study ($n = 2056$) and Avon Longitudinal Study of Parents and Children ($n = 6216$) to study the average causal effect of maternal pregnancy alcohol use on offspring attention deficit hyperactivity disorder symptoms, we proposed 11 maternal SNPs as instruments. We computed bounds assuming subsets of SNPs were jointly valid instruments, for all combinations of SNPs where the MR model was not falsified.

Results: The MR assumptions were violated for all sets with more than 4 SNPs in one cohort and for all sets with more than 2 SNPs in the other. Bounds assuming one SNP was an individually valid instrument barely improved on assumption-free bounds. Bounds tightened as more SNPs were assumed to be jointly valid instruments, and occasionally identified directions of effect, though bounds from different sets varied.

Conclusions: Our results suggest that, when proposing multiple instruments, bounds can contextualise plausible magnitudes and directions of effects. Computing bounds over multiple assumption sets, particularly in large, high-dimensional data, offers a means of triangulating results across different potential sources of bias within a study and may help researchers to better evaluate and emphasise which estimates are compatible with the most plausible assumptions for their specific setting.

KEYWORDS

ALSPAC, bounds, instrumental variables, Mendelian randomisation, MoBa

1 | BACKGROUND

The increasing availability of large-scale genetic data in pregnancy cohorts presents many novel opportunities for pre- and perinatal epidemiologic research. However, it is important to remember that large sample sizes alone will not resolve structural biases resulting from violations of the core assumptions of any given causal inference method. Indeed, as increasing sample sizes allow us to narrow statistical uncertainty, and potentially inform policy and treatment decisions based on the results of observational studies, greater caution is needed to ensure an honest and transparent portrayal of the assumptions upon which any causal inference rests. One setting in which this caution may be especially important is Mendelian randomisation (MR), an approach that leverages genetic variants proposed as instrumental variables¹ to estimate causal effects in the presence of unmeasured confounding. Because MR requires genetic data and larger sample sizes than other causal inference designs,² the viability and appeal of the approach were limited prior to the current era of big data.

When proposing one single nucleotide polymorphism (SNP) as an instrument, MR requires that the SNP is associated with the exposure, does not affect the outcome except through the exposure, and individuals at different levels of the SNP are exchangeable with regards to the counterfactual outcome.³ To obtain a point estimate for the average causal effect of exposure A on outcome Y in the population ($E(Y^a - Y^{a'})$), investigators must additionally make one of a set of possible homogeneity assumptions, described in detail

Synopsis

Study question

Do nonparametric bounds provide useful information in the context of large-scale MR studies of prenatal exposures?

What's already known

Point estimation in MR typically requires strong, unverifiable homogeneity assumptions beyond the core MR assumptions. Bounds, which do not require homogeneity assumptions, are rarely applied in MR.

What this study adds

We computed bounds on the average causal effect of alcohol consumption during pregnancy on offspring ADHD symptoms in two European cohorts, using 11 proposed genetic instruments. Our results suggest that, when proposing multiple instruments, bounds can contextualise plausible magnitudes and directions of effects. Bounds may also highlight the importance of carefully evaluating the assumptions necessary for estimating causal effects, especially as larger, more comprehensive data on perinatal health becomes available.

in the Supplementary Materials and elsewhere.⁴⁻⁷ Unfortunately, these point estimating conditions are often biologically implausible in MR.^{8,9}

In contrast, the bounding of the average causal effect can be conducted under the 3 primary MR conditions alone. Historically, bounding approaches have been unpopular, possibly because bounds based on a single binary proposed instrument are often wide.¹⁰ However, in the high-dimensional data setting in which multiple SNPs could be proposed as instruments, there are under-recognised opportunities. First, because including larger numbers of proposed joint instruments increases instrument strength and allows us leverage multiple assumption sets simultaneously, we might tighten bounds by proposing joint sets of genetic instruments.¹¹⁻¹³ Second, by comparing bounds computed under different assumptions, we might learn more about our reliance on assumptions in informing plausible effect sizes.¹⁴⁻¹⁶

This approach may be especially helpful for MR studies of the effect of pregnancy alcohol consumption on offspring outcomes. While several non-MR studies have found positive associations between maternal pregnancy alcohol use and offspring attention deficit hyperactivity disorder (ADHD),¹⁷⁻¹⁹ these estimated effects may be confounded by other maternal health behaviours. However, because offspring alcohol exposure depends both on the amount of alcohol consumed by the mother and the speed of the mother's alcohol metabolism, most versions of homogeneity assumptions required for point estimation using MR are violated when proposing alcohol dehydrogenase-related SNPs as instruments: The effect of alcohol exposure would likely be heterogeneous across offspring of mothers with different genetic variants.^{8,10} Additionally, because the effect of alcohol exposure is likely heterogeneous across other sociodemographic and health characteristics, homogeneity assumptions are also suspect when proposing non-alcohol dehydrogenase SNPs as instruments.⁸ Here, we demonstrate the use of bounds derived from multiple proposed instruments in an MR study of maternal pregnancy alcohol consumption on offspring ADHD where effect heterogeneity is expected and provide adaptable software for the implementation of the bounds across combinations of proposed instruments. For those more familiar with target trial emulation, we summarise key protocol elements in the Supplementary Materials.^{20,21}

TABLE 1 Prevalence of maternal alcohol use and offspring attention deficit-hyperactivity (ADHD) symptoms in the Avon Longitudinal Study of Parents and Children (ALSPAC) and the Norwegian Mother, Father and Child Study (MoBa).

	ALSPAC	MoBa
	% (n)	% (n)
<i>n</i>	2056	6216
Alcohol Use During 2nd and 3rd Trimesters of Pregnancy		
0 g/week	45.1 (927)	90.1 (5603)
≤32 g/week	23.1 (474)	9.0 (562)
>32 g/week	31.9 (655)	0.8 (51)
Offspring ADHD symptoms	2.3 (47)	2.6 (163)

2 | METHODS

2.1 | Data

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal birth cohort, which aimed to recruit all pregnant women in former Avon County with a due date between 1 April 1991 and 31 December 1992 and continues to follow the offspring. 75.3% of contacted women agreed to participate, resulting in a total of 14,541 pregnancies enrolled during this period. When the oldest children were approximately 7 years old, the study recruited additional eligible children who had not previously participated. The study now includes data on the offspring of 15,454 pregnancies. Further detail is available elsewhere.²²⁻²⁴ The study website contains details on available data through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). We restricted analyses to singleton pairs of self-reported white European origin with complete data on maternal genotype, maternal pregnancy drinking behaviour, and offspring outcomes, resulting in a total analytic sample of 2056 mother-child pairs (Figure S1; Table 1).

The Norwegian Mother, Father and Child Cohort Study (MoBa) is a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health. Participants were recruited from all over Norway from 1999 to 2008. The women consented to participation in 41% of the pregnancies. The cohort now includes 114,500 children, 95,200 mothers and 75,200 fathers. Detailed information is available elsewhere.^{25,26} The current study is based on version 12 of the quality-assured data files released for research in January 2019. For this study, we restricted our sample to singleton pairs with complete data on maternal genetics, maternal pregnancy drinking behaviour and offspring outcomes, resulting in a final analytic sample of 6216 mother-child pairs (Figure S2; Table 1).

2.2 | Measures

2.2.1 | Genetic variants

We selected SNPs based on a recent genome-wide association (GWA) study of alcohol use in UK Biobank.²⁷ Of the 14 SNPs identified at

genome-wide significance in that analysis, we excluded 3 SNPs previously found to be associated with traits that could violate MR assumptions via pleiotropy or were within genes that were associated with such traits.^{28–33} The 11 independent SNPs we thus proposed as instruments were rs145452708, rs193099203, rs11940694, rs29001570, rs3114045, rs140280172, rs9841829, rs35081954, rs9991733, rs149127347 and chr18:72124965. ALSPAC mothers were genotyped using the Illumina human660W-quad and imputed to the 1000 Genome Project. MoBa mothers were genotyped using either Illumina HumanCoreExome or Illumina Global Screening Array, and genotypes were imputed to Haplotype Reference Consortium (HRC) version 1.1. Details of ALSPAC and MoBa genotyping procedures are available in the Supplementary Materials.

In contrast to GWA studies, measurement error of SNPs proposed as instruments will not bias average causal effect estimates of the exposure of interest on the outcome, as long as measurement error of the SNPs is at most differentially associated with the exposure and not with the outcome.⁸ For this reason, we did not exclude proposed instruments with minor allele frequencies under 5% or imputation quality below 0.8. However, assortative mating can violate MR assumptions.³⁴ While Hardy Weinberg equilibrium tests for all SNPs proposed as instruments were conducted as part of the quality control pipeline in both cohorts, these tests may be underpowered to detect small deviations.³⁵ Such deviations could cause large biases in MR. We estimated the correlation between maternal and paternal genotype for each SNP proposed as an instrument in one cohort to identify SNPs which may be particularly vulnerable to this bias (Supplementary Materials).

Because there is incomplete overlap of loci between 1000Genomes and HRC, not all 11 SNPs were available in MoBa. Proxies for unavailable SNPs were selected using LDProxy, based on maximum r^2 .³⁶ Within MoBa, rs145441283 was used as a proxy for rs193099203 and rs1154447 was used as a proxy for rs35081954. Because chr18:72124965 was unavailable in either cohort, rs201288331 was used as proxy in ALSPAC, and rs12955142 was used as a proxy in MoBa.

2.2.2 | Alcohol use

Alcohol use in the second and third trimesters was assessed via postal questionnaire around gestational weeks 18 and 32 in ALSPAC, in which mothers reported their average volume and frequency of alcohol consumption in the last few weeks. In MoBa, mothers reported average volume and frequency of alcohol consumption between gestational weeks 13–24 and after Week 25 via a postal questionnaire at Week 30. Although drinking in pregnancy is not truly a binary process, and mild drinking likely incurs different effects than heavy drinking, the bounding approach used here (described below) requires a binary exposure. For that reason, mothers were categorised as ever drinkers if they reported drinking any amount of alcohol during the second or third trimesters, and never drinkers if they did not report any drinking during either

trimester. Because heavy and moderate drinking were included in the same category, this approach may be vulnerable to bias from poorly defined interventions. To evaluate whether this caused violations of the instrumental inequalities, we applied the instrumental inequalities when grouping alcohol consumption into 4 categories (never drinking, <1 drink per week, 1–2 drinks per week, > 2 drinks per week) and 7 categories (never drinking, <1 drink per week, 1–2 drinks per week, 3–4 drinks per week, 5–6 drinks per week, 7–13 drinks per week, > 13 drinks per week). In secondary analyses, we restricted the study population to compare never drinking and moderate drinking, defined as drinking less than or equal to 32 grams of alcohol per week (approximately 2 drinks per week). Restricting the analytic population in this way can generate selection bias,³⁷ which is why this is not the primary approach.

2.2.3 | ADHD

In ALSPAC, mother-reported ADHD symptoms at age 7 were assessed using the Development and Well-being Assessment.³⁸ In MoBa, mother-reported ADHD symptoms at age 5 were assessed using the Child Behaviour Checklist attention deficit hyperactivity subscale.³⁹ Children with subscale T scores at or above the 98th percentile within the full MoBa cohort (raw score 8, equivalent to the 84th percentile in published norm data) were considered to have ADHD symptoms.³⁹

2.3 | Statistical analysis

When multiple SNPs are believed to be individually valid instruments, several MR models using different subsets of SNPs, and thus slightly different assumptions, are possible. We could conduct MR models separately for each SNP proposed as an instrument. If we were willing to assume several SNPs were individually and jointly valid instruments, we could also conduct MR analyses proposing the set of SNPs as joint instruments. Generally, if we assume faithfulness, any set of independent SNPs that are individually valid instruments will also be jointly valid instruments. However, individually valid instruments will not necessarily be jointly valid instruments if the SNPs are dependent.¹³

Our analysis plan included computing bounds under combinations of assumptions related to the 11 SNPs being proposed as instruments, as described below. Prior to computing any of these bounds, we applied the instrumental inequalities to attempt to falsify each assumption set.^{40,41} In brief, the instrumental inequalities are a set of mathematical constraints implied by the 3 primary MR conditions. Violations of the instrumental inequalities indicate that the observed data distribution is inconsistent with the 3 primary MR conditions, allowing investigators to falsify (but not verify) the MR conditions. Further detail on this approach is available elsewhere.^{9,40,41} In each cohort, we applied the Balke-Pearl instrumental inequalities across all possible combinations of

the SNPs proposed as instruments, including all possible combinations of subsets of the SNPs (e.g. all combinations of 2 SNPs, 3 SNPs, etc.), and to a categorical, unweighted allele score, using code developed previously.⁹ We also applied the Bonet instrumental inequalities to each SNP individually. All sets that violated the instrumental inequalities (e.g. resulted in values greater than 1 for the Balke-Pearl inequalities, or greater than 2 for the Bonet inequalities) were eliminated from further analysis. When multiple SNPs are proposed as joint instruments, the MR model can also be falsified if the bounds calculated using the sets flip, meaning the lower bound is higher than the upper bound. Sets that produced flipped bounds were also removed from the results.

As increasingly large numbers of SNPs are proposed as joint instruments, it is increasingly likely that the MR conditions, and thus the instrumental inequalities, will be violated by chance, rather than by a structural bias in the super-population of interest. These random violations are similar to the concept of 'random confounding' in randomised control trials.⁴² As with random confounding in randomised control trials, if random violations of the MR conditions are present within a sample, an MR analysis in that sample is expected to produce biased effect estimates.⁹ By eliminating all sets that violated the instrumental inequalities, we could eliminate all sets for which the MR conditions were clearly falsified. However, because it is unclear which violations of the inequalities represent structural violations of the MR conditions, as opposed to random violations, the extent to which results of the instrumental inequalities in this study can be generalised to other data sets is unclear. The limited generalisability of the instrumental inequalities is further exacerbated by the possibility of study-specific biases such as selection bias due to missing data.

In the setting of a binary exposure and outcome, bounds on the average causal effect can be calculated using exposure and outcome data alone, without any assumptions.^{5,6} These assumption-free bounds will always have width 1 and always include the null, meaning they cannot identify the direction of effect. Under the MR assumptions, narrower bounds on the average causal effect are possible. When a set of SNPs are assumed to be jointly valid instruments, the set can be combined into a single variable, with levels representing every unique combination of alleles from the included SNPs. This combined variable can then be used to generate bounds using the expression described by Richardson and Robins.¹² To evaluate differences in the bounds across different joint instruments in each cohort, we calculated Richardson–Robins bounds for all possible combinations of the 11 SNPs, including all combinations of subsets of SNPs, that did not violate the instrumental inequalities (Supplementary Materials).

If at least some number k SNPs, but not all 11, were jointly valid instruments, then the average causal effect would lie within the union of the Richardson–Robins bounds computed proposing combinations of k SNPs as joint instruments.¹¹ To explore this, we computed bounds in each cohort assuming only a subset of the 11 SNPs were jointly valid instruments, for all subset sizes where at least some combinations did not violate the instrumental inequalities.

In the context of alcohol–dehydrogenase-related SNPs and prenatal alcohol, the additional homogeneity assumption required for point estimation of the average causal effect in MR is likely invalid. However, in order to explore how conclusions from point estimation and bounding in MR differ, we computed point estimates using two-stage least-squares (Supplementary Materials).

Although MoBa and ALSPAC are relatively ethnically homogenous, residual population stratification may bias our results. We therefore calculated the above-described instrumental inequalities and bounds for each possible combination of the proposed instruments in data weighted using inverse probability weighting to adjust for 10 principal components (Supplementary Materials).³

All analyses were conducted in R version 3.6.1. Adaptable R code for application of the instrumental variable bounds, filtered by the instrumental inequalities, are available in Supplementary Materials.

2.4 | Missing data

Analyses in both cohorts were restricted to individuals with complete data on proposed instruments, exposures and outcomes. While this approach can violate the MR assumptions, it aligns with common practice in MR analyses, allowing for comparison between this approach and previous analyses.

2.5 | Ethics approval

For ALSPAC, informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Ethical approval for ALSPAC was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. The establishment of MoBa and initial data collection was based on a licence from the Norwegian Data Protection Agency and approval from The Regional Committee for Medical and Health Research Ethics. MoBa is now based on regulations related to the Norwegian Health Registry Act. The current study was approved by The Norwegian Regional Committee for Medical and Health Research Ethics (2016/1702).

3 | RESULTS

When comparing any alcohol consumption to no alcohol consumption, in ALSPAC, the instrumental inequalities held for all SNPs individually, 27 combinations of 2 SNPs, 14 combinations of 3 SNPs, two combinations of 4 SNPs, and no combinations of 5 or more SNPs (Figure S3). In MoBa, the instrumental inequalities held for 9 combinations of 2 SNPs and did not hold for any combination of 3 or more SNPs (Figure S4). In addition, the instrumental inequalities failed to hold for 4 SNPs individually in MoBa. A similar amount and pattern of instrumental inequality violations were observed when

comparing moderate alcohol consumption to no alcohol consumption (Figures S5 and S6). Results of the instrumental inequalities were also broadly similar when categorising alcohol consumption into 4 or 7 categories (Figures S7–S10).

In ALSPAC, bounds assuming at least one SNP was an individually valid instrument were very wide (−0.52, 0.47) and barely improved on the assumption-free bounds (−0.54, 0.46). Bounds calculated using each instrument individually were similarly wide (Figure 1). As the number of SNPs assumed to be jointly valid instruments increased, the bounds narrowed substantially, and sometimes fell completely on one side of the null, identifying the direction of effect. However, bounds from different sets of proposed instruments varied substantially, even identifying opposite directions of effect and thus indicating the incompatibility of the assumption sets. With few exceptions, point estimates generally fell within the bounds (Table S1).

In MoBa, bounds were consistent across different assumptions (Figure 2). In all cases, the bounds covered the null. In most cases, the bounds did not differ substantially from the assumption-free bounds (−0.12, 0.88), with the narrowest bounds computed being based on the assumption that two SNPs (rs29001570 and rs9841829) were jointly valid (−0.07, 0.51). In 6 of 16 sets of proposed instruments, point estimates fell outside of the bounds (Table S2).

Bounds computed to estimate the effect of moderate alcohol consumption, rather than any alcohol consumption, followed a similar pattern in both cohorts (Figures 3 and 4, Tables S3 and S4). In ALSPAC, bounds proposing combinations of 3 SNPs narrowed more substantially than the any alcohol models, though bounds still varied substantially and several sets resulted in flipped bounds. Correlation between maternal and paternal genotypes was generally very small (Table S5).

4 | COMMENT

4.1 | Principal Findings

When single SNPs were proposed as instruments, bounds on the average causal effect of both any and moderate prenatal alcohol consumption on offspring ADHD were wide, and were consistent with negative, null and positive effects. However, in ALSPAC, as increasing number of SNPs were assumed to be joint instruments, bounds narrowed and sometimes identified the direction of effect, though bounds varied substantially across different proposed instruments. In MoBa, the instrumental inequalities held for far fewer

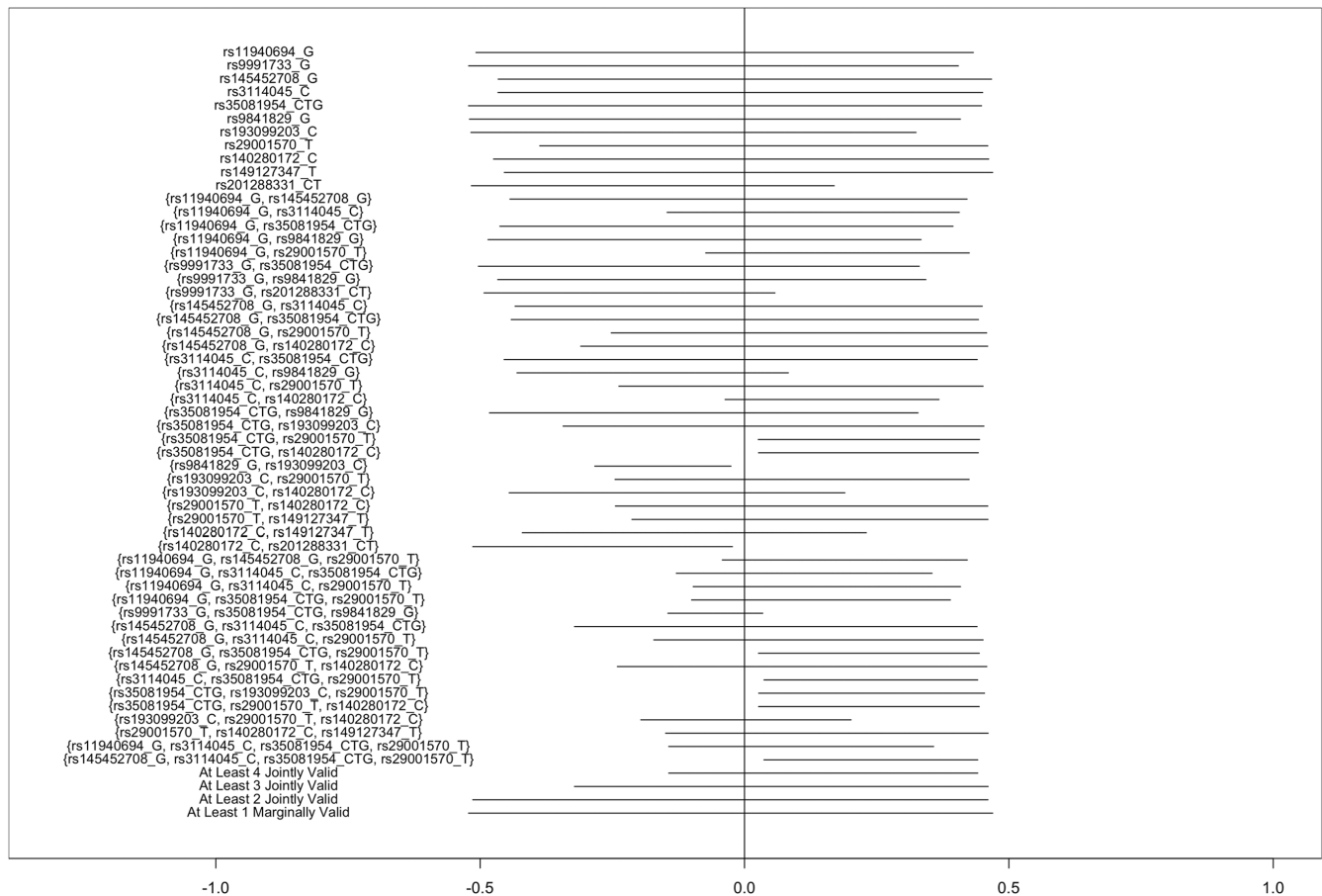


FIGURE 1 Bounds on the average causal effect of any vs no alcohol consumption during the second and third trimesters on offspring attention deficit hyperactivity disorder symptoms in the Avon Longitudinal Study of Parents and Children, inverse-probability weighted for 10 principal components, proposing different sets of SNPs as instruments.

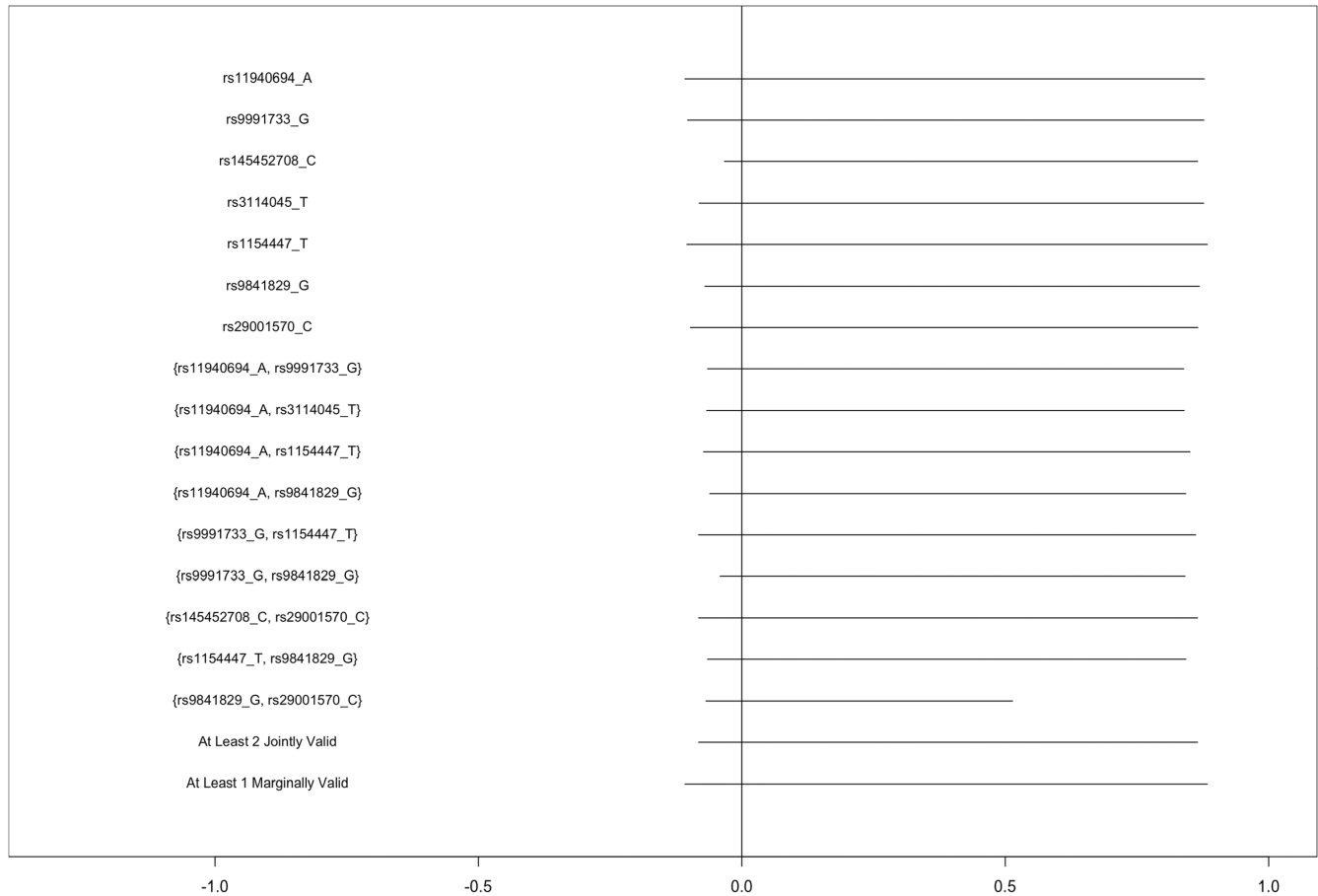


FIGURE 2 Bounds on the average causal effect of any alcohol vs. no alcohol consumption during the second and third trimesters of pregnancy on offspring attention deficit hyperactivity disorder symptoms in the Norwegian Mother, Father and Child Study, inverse probability weighted for 10 principal components, proposing varying combinations of SNPs as instruments.

sets of proposed instruments compared with ALSPAC. Bounds on the average causal effect of moderate and any alcohol consumption on offspring ADHD remained wide and fairly constant across several different sets of assumptions in MoBa.

4.2 | Strengths of the study

Although bounds proposing a single SNP as an instrument barely improved on the assumption-free bounds, the width of the bounds did decrease as we incorporated stronger assumptions. Our ability to evaluate how incorporating stronger assumptions might narrow the bounds was limited by the fact that the strongest assumption sets we considered a priori (that all 11 SNPs were jointly valid instruments) were found to be violated. Nonetheless, bounds in our analysis did narrow as larger numbers of SNPs were proposed as joint instruments, and sometimes identified the direction of effect. Although our analysis was unable to determine a consistent direction of effect across cohorts and proposed instruments, the fact that we observed bounds narrow enough to identify the direction of effect when proposing multiple joint instruments suggests that, in future studies where multiple SNPs are proposed as jointly valid

instruments, bounds may be able to inform decision making without additional point estimating assumptions. This may be especially helpful for contexts, like MR studies of prenatal alcohol exposure, where homogeneity assumptions are implausible.

An advantage of computing bounds over many different assumptions is that such approaches can clarify how different assumptions can change study conclusions.⁴³ In our application, we were only able to identify a direction of effect under the strong assumption that multiple SNPs were jointly valid instruments. Moreover, in ALSPAC, proposing different sets of SNPs as joint instruments resulted in bounds that identified opposite directions of effects. Because such bounds are contradictory, this means that at least one of those two sets of SNPs are invalid joint instruments, although neither violated the instrumental inequalities, which is a finding in and of itself. This kind of incompatibility of assumption sets would have been difficult if not impossible to identify in many MR point estimation approaches but is clearly apparent when bounds are evaluated over several possible assumptions. Computing bounds over many different assumptions about the SNPs proposed as instruments could shift the focus of MR studies towards the question of what assumptions are most plausible, and thus which range of effects we should be most confident in. This property may be enhanced by combining bounding

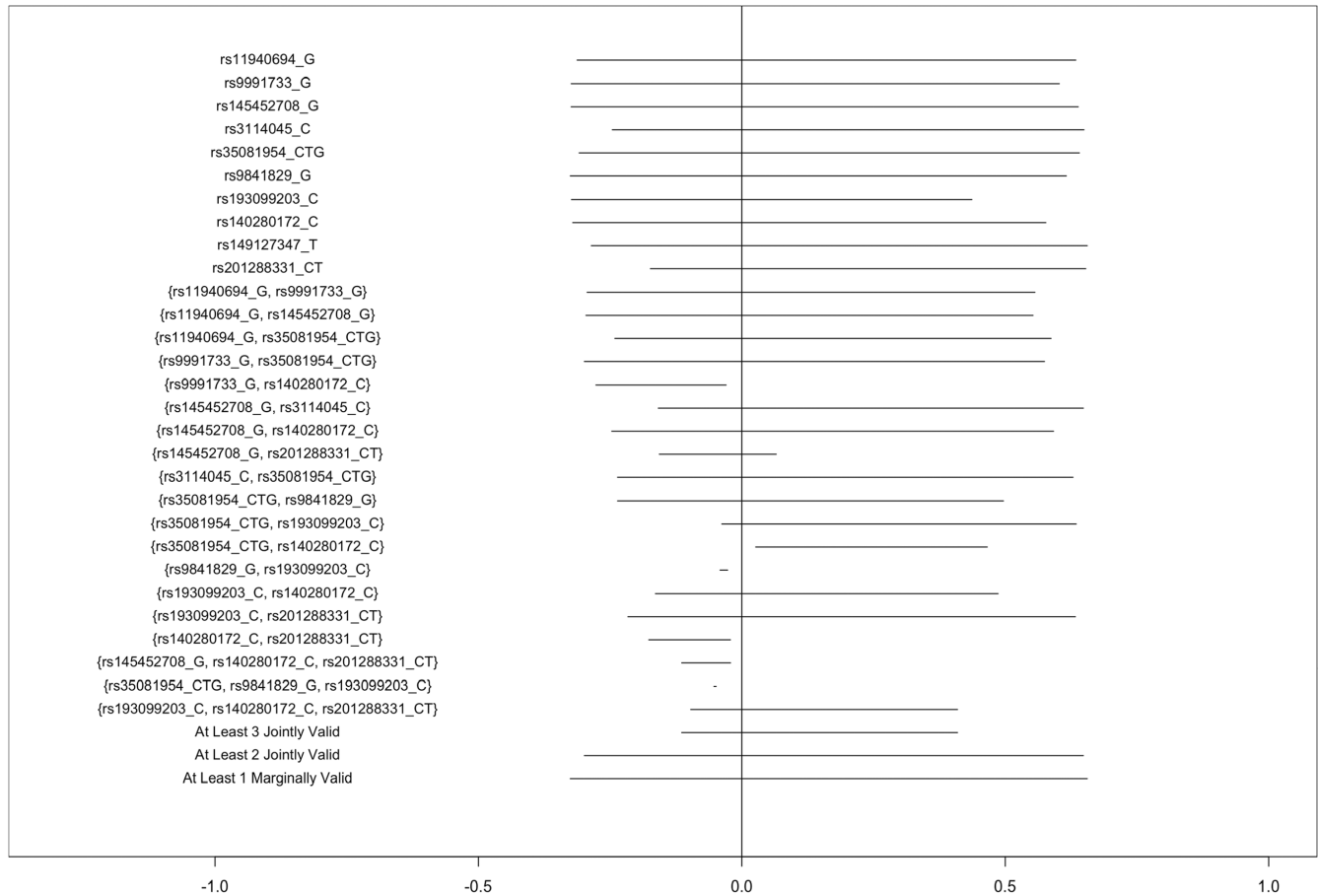


FIGURE 3 Bounds on the average causal effect of moderate (<2 drinks/week) vs no alcohol consumption during the second and third trimesters on offspring attention deficit hyperactivity disorder symptoms in the Avon Longitudinal Study of Parents and Children, inverse-probability weighted for 10 principal components, proposing varying combinations of SNPs as instruments.

with applications of the instrumental inequalities, which could allow for the elimination of analyses based on clearly invalid assumptions.⁹

4.3 | Limitations of the data

Our results showed that at least 7 of the 11 SNPs in our analysis could not be valid instruments in ALSPAC, and at least 9 of the 11 could not be valid instruments in MoBa. This is surprising, as the full set of proposed instruments contained 4 SNPs in alcohol dehydrogenase regions, whose relationship to alcohol consumption is relatively well understood. This detected bias could have resulted from several different causes (some of which are detailed in the Supplementary Materials),⁴⁴ but indicates that MR studies of prenatal alcohol exposure may be more vulnerable to bias than was previously understood and should be viewed with caution.

4.4 | Interpretation

In general, it is important to recognise both that the instrumental inequalities are a falsification test only and that violations of the

instrumental inequalities when proposing sets of joint instruments do not indicate which SNPs in the set are invalid instruments or the source of the bias. For these reasons, future MR analyses may use instrumental inequalities in concert with other falsification techniques and subject matter knowledge to evaluate the validity of their assumptions. In this example, our results suggest further investigation is needed to clarify how maternal alcohol-related SNPs impact offspring behavioural health.

The variation in the bounds across assumption sets also illustrates how strongly point estimation in MR relies on the homogeneity assumptions (described in the Supplementary Materials). Even under the strongest unfalsified assumption sets, bounds often covered a moderately large range of effect sizes, meaning point estimation under those sets would still depend heavily on the homogeneity assumptions. Under weaker sets of assumptions, like proposing a single SNP as an instrument, the conclusions of MR studies using point estimation would be informed almost entirely by those additional homogeneity assumptions. This suggests that greater attention should be paid to evaluating the validity of point-estimating assumptions in MR. In our application, point estimates sometimes fell outside the bounds, indicating a violation of the point-identifying assumptions. These sets included SNPs inside and outside of alcohol

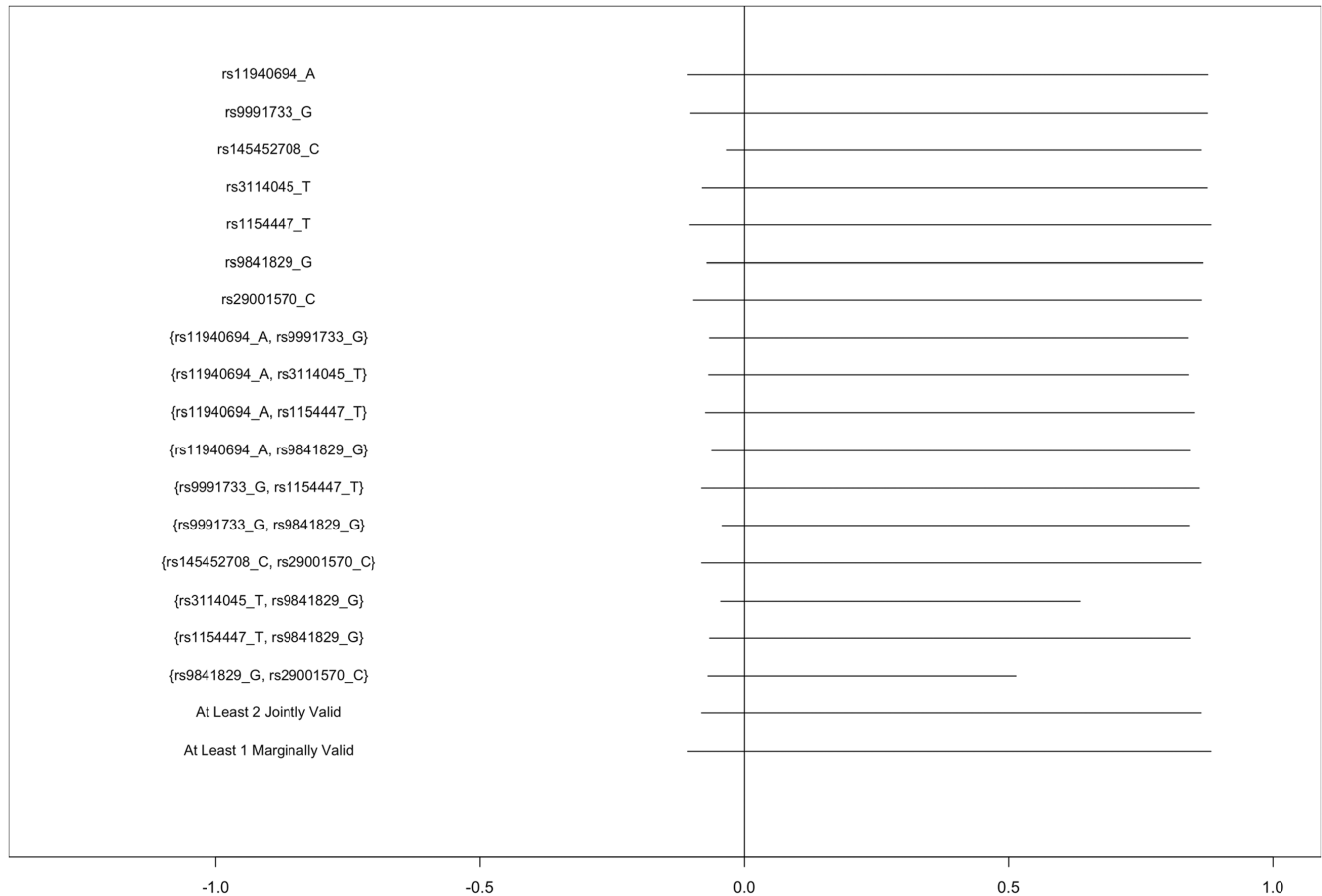


FIGURE 4 Bounds on the average causal effect of moderate alcohol consumption moderate (<2 drinks/week) vs no alcohol consumption during the second and third trimesters on offspring attention deficit hyperactivity disorder symptoms in the Norwegian Mother, Father and Child Study, inverse-probability weighted for 10 principal components, proposing varying combinations of SNPs as instruments.

dehydrogenase regions. While violations of homogeneity were expected in our context, this suggests the resulting bias was severe, and future MR studies might benefit from closer evaluation of the plausibility of the point estimating assumptions.

Even in settings where both the primary MR assumptions and the additional point estimating assumptions are plausible, the presentation of the bounds alongside point estimates could help readers and investigators to understand how strongly MR studies depend on assumptions. This is true even, and perhaps especially when the bounds are wide. Several studies have called for the presentation of bounds in observational studies, particularly for instrumental variable models like MR.^{14–16,43} Robins and Greenland noted that ‘wide bounds make clear the degree to which public health decisions are dependent on merging the data with strong prior beliefs.’¹⁵ Incorporating bounds into MR practice would clarify the amount of information present in the data alone, and the need for critical evaluation of assumptions within each study’s unique context.

Further research is needed to extend bounding approaches for instrumental variables and MR in several ways, including but not limited to extensions for: estimation procedures incorporating sampling variability^{14,45}; time-varying interventions^{46–48}; conditional

instrumental variables incorporating measured covariates⁸; non-binary exposures^{49,50}; and two-sample approaches.⁵¹ Though this list is not exhaustive, we believe it represents priorities for maximising the usefulness and applicability of bounding in MR.

5 | CONCLUSIONS

The availability of data on increasingly large, genetically informed pregnancy cohorts opens a number of new avenues for perinatal epidemiology, including allowing for larger and thus more potentially informative MR studies of prenatal exposures and outcomes. However, larger sample sizes alone will not prevent bias resulting from structural violations of the required MR assumptions or any other causal inference approach. MR studies frequently propose large numbers of SNPs as joint instruments and thus make large numbers of assumptions about the joint validity of those proposed instruments. Adding to the growing arsenal of sensitivity analyses, bounding may allow researchers to leverage these assumptions to make meaningful conclusions about effects without additional, potentially implausible homogeneity assumptions. Even when homogeneity assumptions are biologically plausible, estimating bounds across different combinations of

proposed instruments may allow investigators to better evaluate the dependence of their conclusions on those assumptions.

More broadly, the substantial variation seen across bounds computed using different proposed instruments in our study highlights how strongly the result of any causal analysis depends on the assumptions made by investigators, and thus how important it is to interrogate, be transparent about, and remain humble when making those assumptions. Particularly as sample sizes and/or data dimensionality increase, incorporating bounds into a larger range of epidemiologic studies could help shift conversations around causal effects away from one or few point estimates and towards questions about what ranges of estimates are consistent with assumptions readers and investigators feel most confident in.

ACKNOWLEDGEMENTS

We thank Hannah Sallis and Laurie Hannigan for their data analytic support. We thank Miguel Hernán for his helpful feedback on earlier drafts of this work. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (NRC) (#229624). We also thank the NORMENT Centre for providing genotype data, funded by NRC (#223273), South East Norway Health Authority and KG Jebsen Stiftelsen. Further we thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTIONPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, NRC, the Novo Nordisk Foundation, the University of Bergen and the Western Norway health Authorities (Helse Vest). The Norwegian Mother and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. This work was performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT). (tsddrift@usit.uio.no). We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

FUNDING INFORMATION

This work was partly supported by the South-Eastern Norway Regional Health Authority (2018059) and the Norwegian Research Council (274611). This project is supported by an innovation programme under the Marie Skłodowska-Curie grant agreement no. 721567. Dr Swanson is further supported by a NWO/ZonMW Veni Grant (91617066). H. Tiemeier is supported by a grant of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organisation for Scientific Research (NWO grant No. 024.001.003, Consortium on Individual Development). MR Munafò

and L Zuccolo are part of the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/7, MC_UU_00011/1). The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors, and all authors will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website. Maternal ADH SNP genotyping was funded by MRC grant number G0902144. ALSPAC GWAS data was generated by Sample Logistics and Genotype Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. The design of questions on parental alcohol consumption was funded by a grant from the National Institute of Alcohol Abuse and Alcoholism to Dr. Ruth Little. This research was also supported by the NIHR Bristol Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health and Social Care.

AUTHOR CONTRIBUTIONS

EWD and SAS conceptualized the study and designed the analytic plan. EWD performed data analysis and wrote the initial draft of the manuscript. SAS, HT, AH, LZ, and MRM supervised the analysis and revised the manuscript. MRM, OAA, and PRN provided the data and gave critical feedback on the manuscript.

CONFLICT OF INTEREST

The authors declared no conflict of interest for this article.

DATA AVAILABILITY STATEMENT

The ALSPAC data management plan (<http://www.bristol.ac.uk/alspac/researchers/data-access/documents/alspac-data-management-plan.pdf>) describes in detail the policy on data sharing, which is through a system of managed open access. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool and reference the following webpage: <http://www.bristol.ac.uk/alspac/researchers/our-data/>. MoBa data are used by researchers and research groups at both the Norwegian Institute of Public Health and other research institutions nationally and internationally. The research must adhere to the aims of MoBa and the participants' given consent. All use of data and biological material from MoBa is subject to Norwegian legislation. More information can be found on the study website (<https://www.fhi.no/en/studies/moba/for-forskere-artikler/research-and-data-access/>).

REFERENCES

1. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res.* 2007;16:309-330.
2. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32:1-22.

3. Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC; 2018.
4. Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance. *J Am Stat Assoc*. 1997;92:1171-1176.
5. Manski CF. Nonparametric bounds on treatment effects. *Am Econ Rev*. 1990;80:319-323.
6. Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Bailey L, eds. *Health Service Research Methodology: A Focus on AIDS*. US Public Health Service; 1989:113-159.
7. Tchetgen Tchetgen EJ, Sun B, Walter S. The GENIUS approach to robust mendelian randomization inference. arXiv Preprint arXiv:170907779. 2017.
8. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17:360-372.
9. Diemer EW, Labrecque J, Tiemeier H, Swanson SA. Application of the instrumental inequalities to a mendelian randomization study with multiple proposed instruments. *Epidemiology*. 2020;31:65-74.
10. Swanson SA, Hernán MA. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology*. 2013;24:370-374.
11. Swanson SA. Commentary: can we see the forest for the IVs? Mendelian randomization studies with multiple genetic variants. *Epidemiology*. 2017;28:43-46.
12. Richardson TS, Robins JM. ACE bounds; SEMs with equilibrium conditions. *Statistical Science*. 2014;29:363-366.
13. Sheehan NA, Didelez V. Epidemiology, genetic epidemiology and mendelian randomisation: more need than ever to attend to detail. *Hum Genet*. 2020;139:121-136.
14. Swanson SA, Hernán MA, Miller M, Robins JM, Richardson TS. Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes. *J Am Stat Assoc*. 2018;113:933-947.
15. Robins JM, Greenland S. Identification of causal effects using instrumental variables: comment. *J Am Stat Assoc*. 1996;91:456-458.
16. Cole SR, Hudgens MG, Edwards JK, et al. Nonparametric bounds for the risk function. *Am J Epidemiol*. 2019;188:632-636.
17. Han J-Y, Kwon H-J, Ha M, et al. The effects of prenatal exposure to alcohol and environmental tobacco smoke on risk for ADHD: a large population-based study. *Psychiatry Res*. 2015;225:164-168.
18. Pagnin D, Grecco MLZ, Furtado EF. Prenatal alcohol use as a risk for attention-deficit/hyperactivity disorder. *Eur Arch Psychiatry Clin Neurosci*. 2019;269:681-687.
19. Linnet KM, Dalsgaard S, Obel C, et al. Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: review of the current evidence. *Am J Psychiatry*. 2003;160:1028-1040.
20. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183:758-764.
21. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70-75.
22. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort profile: the Avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013;42:97-110.
23. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Research*. 2019;4:51.
24. Boyd A, Golding J, Macleod J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon longitudinal study of parents and children. *Int J Epidemiol*. 2013;42:111-127.
25. Magnus P, Birke C, Vejrup K, et al. Cohort profile update: the Norwegian mother and child cohort study (MoBa). *Int J Epidemiol*. 2016;45:382-388.
26. Paltiel L, Anita H, Skjerden T, et al. The biobank of the Norwegian Mother and Child Cohort Study—present status. *Norsk Epidemiologi*. 2014;24:29-35.
27. Clarke T-K, Adams MJ, Davies G, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK biobank (N=112 117). *Mol Psychiatry*. 2017;22:1376-1384.
28. Linnér RK, Biroli P, Kong E, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet*. 2019;51:245-257.
29. Chambers JC, Elliott P, Zabaneh D, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet*. 2008;40:716-718.
30. Kanai M, Akiyama M, Takahashi A, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet*. 2018;50:390-400.
31. Zhong VW, Kuang A, Danning RD, et al. A genome-wide association study of bitter and sweet beverage consumption. *Hum Mol Genet*. 2019;28:2449-2457.
32. Strawbridge RJ, Ward J, Cullen B, et al. Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK biobank cohort. *Transl Psychiatry*. 2018;8:1-11.
33. Brazel DM, Jiang Y, Hughey JM, et al. Exome chip meta-analysis fine maps causal variants and elucidates the genetic architecture of rare coding variants in smoking and alcohol use. *Biol Psychiatry*. 2019;85:946-955.
34. Hartwig FP, Davies NM, Davey SG. Bias in Mendelian randomization due to assortative mating. *Genet Epidemiol*. 2018;42:608-620.
35. Salanti G, Amountza G, Ntzani EE, Ioannidis JPA. Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *Eur J Hum Genet*. 2005;13:840-848.
36. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555-3557.
37. Swanson SA, Robins JM, Miller M, Hernán MA. Selecting on treatment: a pervasive form of bias in instrumental variable analyses. *Am J Epidemiol*. 2015;181:191-197.
38. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry*. 2000;41:645-655.
39. Achenbach TM, Rescorla LA. *Manual for the ASEBA Preschool Forms and Profiles*. University of Vermont, Research center for children, youth ...; 2000.
40. Pearl J. On the testability of causal models with latent and instrumental variables. *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc.; 1995:435-443.
41. Bonet B. Instrumentality tests revisited. *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc.; 2001:48-55.
42. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *Eur J Epidemiol*. 2015;30:1101-1110.
43. Swanson SA, Holme Ø, Løberg M, et al. Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening. *Trials*. 2015;16:541.

44. Diemer EW, Labrecque JA, Neumann A, Tiemeier H, Swanson SA. Mendelian randomisation approaches to the study of pre-natal exposures: a systematic review. *Paediatr Perinat Epidemiol*. 2021;35:130-142.
45. Tamer E. Partial identification in econometrics. *Annu Rev Econ*. 2010;2:167-195.
46. Labrecque JA, Swanson SA. Interpretation and potential biases of mendelian randomization estimates with time-varying exposures. *Am J Epidemiol*. 2019;188:231-238.
47. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat-Theory Methods*. 1994;23:2379-2412.
48. Robins JM. Structural nested failure time models. *Wiley StatsRef: Statistics Reference Online*; 2014. doi:[10.1002/9781118445112.stat06059](https://doi.org/10.1002/9781118445112.stat06059)
49. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol*. 2018;33:947-952.
50. VanderWeele TJ, Tchetgen EJT, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology*. 2014;25:427-435.
51. Lawlor DA. Commentary: two-sample mendelian randomization: opportunities and challenges. *Int J Epidemiol*. 2016;45:908-915.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Diemer EW, Havdahl A, Andreassen OA, et al. Bounding the average causal effect in Mendelian randomisation studies with multiple proposed instruments: An application to prenatal alcohol exposure and attention deficit hyperactivity disorder. *Paediatr Perinat Epidemiol*. 2022;00:1-12. doi:[10.1111/ppe.12951](https://doi.org/10.1111/ppe.12951)