Journal of Developmental Origins of Health and Disease

www.cambridge.org/doh

Challenges in using data on fathers/partners to study prenatal exposures and offspring health

Kayleigh E. Easey[1,2,5,*], Apostolos Gkatzionis[1,2,*], Louise A.C. Millard[1,2], Kate Tilling[1,2], Deborah A. Lawlor[1,2,4] and Gemma C. Sharp[1,2,3]

[1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; [2]School of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; [3]School of Psychology, University of Exeter, Exeter, UK; [4]Bristol National Institute for Health Research (NIHR) Biomedical Research Centre, Bristol, UK and [5]School of Psychological Science, University of Bristol, Bristol, UK

Original Article

Cite this article: Easey KE, Gkatzionis A, Millard LAC, Tilling K, Lawlor DA, and Sharp GC. (2024) Challenges in using data on fathers/partners to study prenatal exposures and offspring health. *Journal of Developmental Origins of Health and Disease* **15**: e25, 1–9. doi: 10.1017/S2040174424000199

Received: 25 August 2023
Revised: 14 May 2024
Accepted: 1 June 2024

Keywords:
Father; partner; pregnancy; selection bias; Avon Longitudinal Study of Parents and Children; Born in Bradford; Millennium Cohort Study

Corresponding author:
Kayleigh E. Easey;
Email: Kayleigh.easey@bristol.ac.uk

*Kayleigh E. Easey and Apostolos Gkatzionis contributed equally.

Abstract

Paternal exposures (and other non-maternal factors) around pregnancy could have important effects on offspring health. One challenge is that data on partners are usually from a subgroup of mothers with data, potentially introducing selection bias, limiting generalisability of findings. We aimed to investigate the potential for selection bias in studies using partner data.

We characterise availability of data on father/partner and mother health behaviours (smoking, alcohol, caffeine and physical activity) around pregnancy from three UK cohort studies: the Avon Longitudinal Study of Parents and Children (ALSPAC), Born in Bradford and the Millennium Cohort Study. We assess the extent of sample selection by comparing characteristics of families where fathers/partners do and do not participate. Using the association of parental smoking during pregnancy and child birthweight as an example, we perform simulations to investigate the extent to which missing father/partner data may induce bias in analyses conducted only in families with participating fathers/partners.

In all cohorts, father/partner data were less detailed and collected at fewer timepoints than mothers. Partners with a lower socio-economic position were less likely to participate. In simulations based on ALSPAC data, there was little evidence of selection bias in associations of maternal smoking with birthweight, and bias for father/partner smoking was relatively small. Missing partner data can induce selection bias. In our example analyses of the effect of parental smoking on offspring birthweight, the bias had a relatively small impact. In practice, the impact of selection bias will depend on both the analysis model and the selection mechanism.

Introduction

The Developmental Origins of Health and Disease (DOHaD) hypothesis posits that environmental factors experienced in early life can have long-lasting influences on health disease risk. Much of DOHaD research is conducted on prospective, longitudinal (birth) cohort studies, with a strong focus on the effects of the intrauterine environment influenced by maternal exposures during pregnancy.[1-4] These maternal pregnancy exposures are most commonly hypothesised to affect child health through direct influences on the developing fetus *in utero* (although transmission of epigenetic modifications via the maternal germline have also been studied.)[5]

Increasing evidence suggests that father/partner exposures (and other non-maternal factors) around pregnancy could also have important effects on offspring health.[6-8] There are several potential mechanisms through which this can occur—either directly (e.g., for biological fathers, via germline transmission of epigenetic modifications) or indirectly via the mother (e.g., for any partner, through mothers passive smoking, or partners support that helps the mother quit smoking or drinking alcohol)[6,9–11] (Figure 1). However, for research measuring paternal exposures around pregnancy, there is a lower breadth (fewer variables measured) and depth (simpler categories, collected less frequently) of available data. This can contribute to a lower quantity of published papers measuring paternal pregnancy effects.[2,8] Smoking in pregnancy for example, is a key maternal exposure known to cause lower birthweight in offspring,[12,13] yet studies seldom measure fathers/partners smoking behaviour in the same detail. Furthermore, in general, women are more likely than men to participate in research when invited, and in follow-up after pregnancy mothers are more likely to be the point of contact and respond to questionnaires about their children or bring them to research clinics collecting measurements and biological samples.[17,18] This results in an imbalance in availability and completeness of data on mothers compared to partners.

The focus on maternal exposures over and above partner exposures in pregnancy could be argued to be due to a lack of availability of partner data in human cohort studies. However, a review of studies concerned with DOHaD, showed that lack of research interest in partner

© The Author(s), 2024. Published by Cambridge University Press in association with The International Society for Developmental Origins of Health and Disease (DOHaD). This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

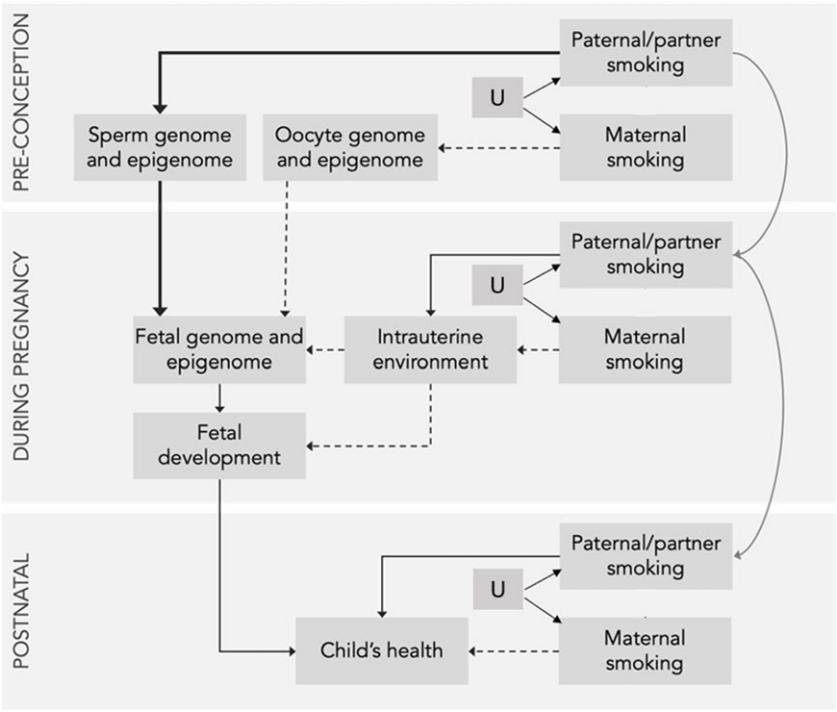


Figure 1. Potential pathways through which father/partner smoking could affect child health. Paternal/partner effects not via the maternal exposure are shown with solid arrows, and effects via maternal exposure are shown with dashed arrows. The thicker solid arrows highlight pathways that are only relevant to fathers/partners who are genetically related to the child (i.e., biological fathers). All other pathways are relevant to all partners. In these models there are unmeasured confounders, of which there will many and therefore represented by U. For simplicity, arrows between maternal smoking at different time periods are omitted, and likewise for the confounder.

effects extends to animal studies too: only 10% of animal studies included partner measures, despite male animals being widely available for research.[2] In some situations, there may be good reason for assuming maternal exposures are likely to have a stronger effect, but without exploring the impact partner exposures may have on offspring outcomes, we are ignoring an important pathway, as well as potentially reinforcing assumptions that it is only maternal behaviours during pregnancy that are important.[1]

The paucity of available partner data has several implications for studies investigating the effects of parental exposures on offspring outcomes. First, it limits both the range of partner exposures in the prenatal period that can be studied and the ability to harmonise measures between parents and cohorts. Second, differences in partner characteristics (e.g., socio-economic position or health behaviours) between families with and without participating partners can induce selection bias in applied analyses using data only on families with participating partners. And third, in addition to investigating the impact partner prenatal behaviours have on child health, the collection of quality data on partners would improve analyses of maternal effects by being able to conduct negative control analyses using a partner exposure. Partner negative control studies are a method used to explore if maternal effects may be biased by residual confounding,[14-16] and account for and block pathways of association from the partner exposure. Unfortunately, the lack of good-quality partner data in some cohorts can inhibit researchers who may wish to conduct such negative control studies. The aim of this paper is to demonstrate the lack of paternal/partner data in many existing cohorts and to assess the impact of selection bias caused by this lack of data on studies investigating the effects of parental exposures on offspring outcomes. To achieve this, we use data from three UK cohort studies: the Avon Longitudinal Study of Parents and Children (ALSPAC), Born in Bradford (BiB) and the Millennium Cohort Study (MCS). We describe the availability of partner data in these three cohorts and illustrate the potential for selection bias due to missing partner data in both simulations and real-data analyses.

Methods

In this section, we first describe the three cohorts used in our analyses — ALSPAC, BiB and MCS —and discuss the collection of partner data in these cohorts. We then conduct three main analyses. First, we summarise available mother and partner data on four health behaviours (smoking, alcohol, caffeine, physical activity) around pregnancy in each of the three cohorts. Second, using ALSPAC data, we conduct simulations exploring the extent to which missing partner data (hence selection into analytic subsamples) can bias estimates, using the association of parental smoking during pregnancy and child birthweight as an example. And third, using data from all three cohorts, we estimate associations of maternal smoking with offspring birthweight in families with and without participating partners, to explore the 'real life' impact of partner sample selection on estimates of maternal effects.

For simplicity throughout the remainder of the paper we use the term 'partner' to include biological fathers and all other partners.

Cohorts

ALSPAC, BiB, and MCS are UK-based prospective, ongoing multigenerational cohorts that recruited women and their partners during pregnancy (ALSPAC and BiB) or via household contact where there was a registered birth within a one-year set period (MCS). They include different populations (by geography and year of recruitment) with different data collection and follow-up protocols, as described below, enabling us to investigate how the impact of selection differs across them. They also used different approaches to obtain partner data, and therefore understanding the selection bias they may include would have a large impact for future research. We have therefore only included these three UK-based cohorts as exemplars of three highly contrasting partner recruitment and follow-up strategies.

https://doi.org/10.1017/S2040174424000199 Published online by Cambridge University Press

Avon longitudinal study of parents and children (ALSPAC)

ALSPAC recruited pregnant women residing in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992.[19,20] The initial number of pregnancies enrolled was 14,541 pregnancies resulting in 14,062 live births, and 13,988 children who were alive at 1 year of age. Participants have been regularly followed up through clinic visits and questionnaires. Ethics approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. The study website (http://www.bristol.ac.uk/alspac/researchers/our-data/) contains details of all the data that is available through a fully searchable data dictionary and variable search tool. For the purposes of this study, we selected the first child from each cohort mother ($n = 14,472$). Pregnant mothers were posted questionnaires at 18 weeks' gestation to give to their partner to answer. If partners agreed, they were subsequently enrolled in the study. If mothers responded that they either had no partner, or that her partner did not want to take part then no further partner questionnaires were sent. Some information on partners was also collected by maternal report. For further details on the cohort profile, representativeness, and phases of recruitment, see articles by Boyd and colleagues[20]; Fraser and colleagues[19] and Northstone and colleagues.[21]

Born in Bradford (BiB)

BiB recruited pregnant women residing in the town of Bradford, in the north of England (UK) between 2007 and 2011.[22] After being provided with initial study information at their first antenatal clinic appointment, pregnant women were largely recruited at a routine oral glucose tolerance test (OGTT) at 24–28 weeks' gestation that all pregnant women in Bradford at the time of the study were invited to, unless they had pre-existing diabetes, which is very unusual and contrary to NICE guidelines.[23] Around 80% of pregnant women in Bradford (without pre-existing diabetes) attended for an OGTT, and about 80% of these participated in BiB. This resulted in the recruitment of 12,453 women with 13,776 pregnancies and 13,740 live births.[22] Ethical approval for BiB was obtained from the Bradford NHS Research Ethics Committee. We excluded mothers who had not completed the mother questionnaire at recruitment (i.e., those who completed it later in pregnancy or postnatally), giving a sample size of 11,395. For the purposes of this study, we randomly selected one child from each cohort mother. Mothers were not asked to report partners behaviours. Partner data was collected directly via self-report, through invitations to participate if they attended the OGTT or subsequent antenatal appointments with the pregnant study mother, or postnatally when a member of the study team visited the family home during follow-up visits.

Millennium cohort study (MCS)

MCS recruited parents of children recorded as a live birth in the United Kingdom between 1st September 2000 and 31st August 2001 (for England and Wales), and between 24th November 2000 and 11th January 2002 (for Scotland and Northern Ireland).[24] To be eligible, the children had to be alive and resident in the UK at age 9 months, resulting in ~ 19,000 children. Eligible children were identified using government child benefit records, which has almost universal coverage. Certain subgroups were intentionally over-sampled (designed for a disproportionate representation of families living in child poverty: children living in disadvantaged areas, children of ethnic minority backgrounds, and children growing up in the smaller nations of the UK), to be nationally representative of the UK population. In the first sweep, 18,827 children from 18,552 families were recruited. Ethical approval for the study was obtained from an NHS Research Ethics Committee and informed consent was obtained from parents for their children's participation. For the purposes of this study, we excluded 55 children where the mother was not the biological mother, and randomly selected one child from each pregnancy ($n = 18,241$).

Within a household both mothers and partners were invited to take part via interview, at which information on non-resident parents were also collected, via the parent who was resident. The first sweep interviewed both mothers and (where resident) fathers/other parents of infants included in the sample. Non-resident parents were invited to take part within the later sweeps.

Extraction and harmonisation of data

Paternity and maternity status

In ALSPAC and BiB, all cohort mothers were recruited during pregnancy and were biologically related to their children. In ALSPAC, mothers were asked whether the partner they had identified to be invited to the study was the biological father. In BiB, cohort mothers/partner were not asked whether the partner was the biological father. In MCS, co-resident partners in the household were asked to report their relationship to the cohort child, including whether they were the child's biological parent. We excluded families where the household main respondent reported that they were not the biological mother.

Definition of partner participation

For each cohort, we created a binary variable to describe the participation of the partners of cohort mothers. Partners were defined as 'participating' if they had provided any data via self-report questionnaire or interview in the initial stages of the studies. In ALSPAC, participating partners had to have completed and returned at least one item from the first three questionnaires given to partners around pregnancy, after enrolment through the pregnant mother. In BiB, participating partners had to have returned the baseline questionnaire, sent to partners during or shortly after the study pregnancy. In MCS, participating partners had to have been interviewed during the first sweep interview.

Partner and maternal health behaviours

For each cohort, we described availability of measures of parental smoking, alcohol, caffeine and physical activity in the prenatal period, to demonstrate the disparities in available data across specific health behaviours measured in the included cohorts. Exposure timepoints of interest were the three months prior to pregnancy (preconception), any time during pregnancy, and in each of the three trimesters separately. For smoking, we generated binary (any/none) variables to describe smoking behaviour of both mothers and partners for each timepoint. More details are provided in the supplementary methods

Other partner, maternal and birth characteristics

Across each cohort, we derived and harmonised (where possible) variables describing parent and birth characteristics. Self-reported measures were maternal age in years at conception (BiB) or delivery (ALSPAC and MCS); partner age during pregnancy (BiB) or delivery (ALSPAC and MCS); highest level of education; mother

https://doi.org/10.1017/S2040174424000199 Published online by Cambridge University Press

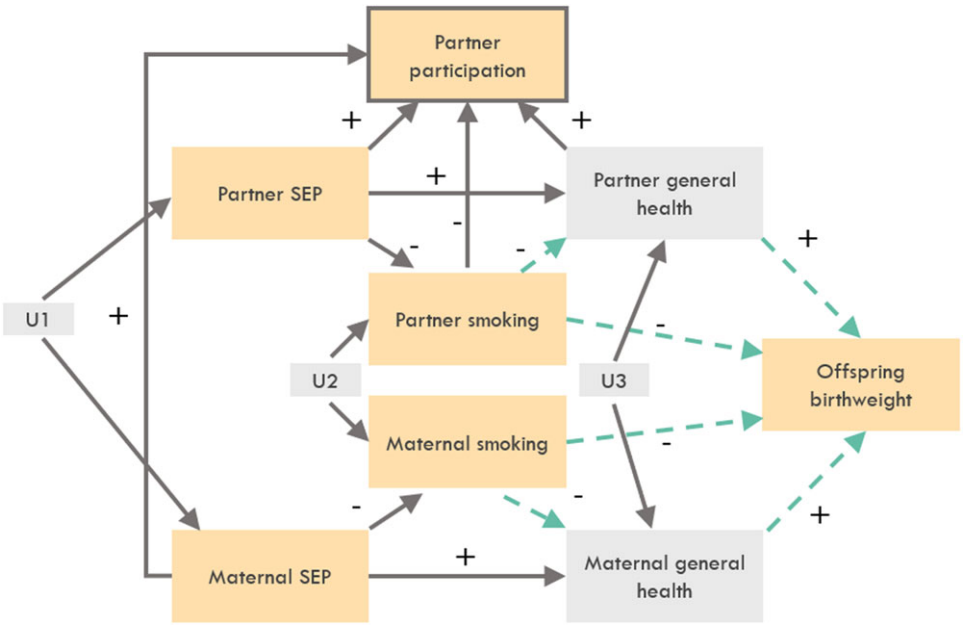


Figure 2. Directed acyclic graph illustrating how missing partner data (incomplete partner participation) might introduce selection bias. Grey shading indicates variables that are not usually measured in real life studies (but were simulated). Dashed green arrows indicate the main effects being estimated and other arrows indicate causal effects. A box is drawn around partner participation to indicate that it is conditioned on through selection.

and partner ethnicity (White European or Black/Asian/mixed/other); maternal pre-pregnancy body mass index (BMI). Mother-reported co-habitation status (partner lives with mother during pregnancy vs not), marital status (married vs not married). Mother-report or birth record of baby's birthweight (grams); gestational age at delivery (weeks). To compare differences in participation across different characteristics on the same scale, we categorised continuous and ordinal measures: parental age (young ≤ 21 years; high ≥ 35 years); maternal pre-pregnancy obesity (low BMI<30 kg/m^2; high BMI ≥ 30) baby's birthweight (low<2.5 kg; high>4.5 kg); preterm birth (< 37 weeks vs ≥ 37 weeks); parental education (university degree or higher vs any lower qualifications/no qualifications)

Because partner data is not generally available for partners that do not participate, we included mother and birth characteristics (maternal age, obesity, ethnicity, education, co-habitation status, marital status, birthweight, gestational age at birth).

Descriptive analysis of cohort characteristics and sample selection

To characterise the availability of data on partner and mother health behaviours in the prenatal period, we summarised the availability (yes/no) of our health behaviours of interest and calculated the percentage of partners participating in each cohort.

As we mentioned earlier, differences in socio-economic position (SEP) and parental health behaviours between families with participating partners and those without could introduce selection bias. To assess the potential for sample selection in these cohorts we used logistic regression to estimate the univariable association of family characteristics in pregnancy with partner participation. Where partner data had been provided by the cohort mother, we also compared partner participation across mother-reported partner's age, ethnicity, education, and paternity status.

Simulation study to explore the magnitude and direction of bias due to sample attrition, using the effects of maternal and partner smoking during pregnancy on birthweight as an example

We conducted a simulation study to assess the impact of selection bias, using the association of maternal and partner smoking during pregnancy with child birthweight as an example. The directed acyclic graph (DAG) in Figure 2 shows our assumptions about the relationship between parental smoking (exposure) and offspring birthweight (outcome), as well as other variables that may influence partner participation (parental SEP and general health). The DAG illustrates how missing partner data might introduce selection bias in studies of the effects of partner or maternal smoking on child birthweight.

We based our simulations on the ALSPAC dataset, since there was more information on partner variables in ALSPAC than the other two cohorts (Figure 3). Our simulations were designed using the ADEMP framework,[25] a commonly used set of guidelines for performing simulation studies. It consists of five steps: (a) state the Aim of the simulation, (b) describe the Data-generating model, (c) specify the Estimand, i.e. the parameter that we want to estimate, (d) describe the Methods used to estimate it and (e) select the Performance measures to be used to report the results. The ADEMP formulation of our simulation study was as follows:

Aim: To assess the impact of selection bias in an ALSPAC-based analysis of the association between parents' smoking and offspring birthweight.

Data-generating mechanism: We generated data on parental smoking, SEP and general health, as well as offspring birthweight, for 15,000 individuals, according to the DAG in Figure 2. Associations between these variables were estimated in ALSPAC and used to parameterise the simulations. Parental 'general health' could not be measured directly in ALSPAC, so for the purpose of this simulation we used BMI as a proxy.

To explore the impact of selection bias, we varied the effects of partner SEP, smoking and general health on partner participation. We generated eleven simulation scenarios, summarised in Supplementary Table 1. In Scenario 1, we set the effects of partner health and smoking on partner participation equal to the observed associations of maternal BMI and smoking with partner participation in ALSPAC (0 and -0.81, respectively) and estimated the association of partner SEP with participation using data reported by mothers about their partners. In Scenarios 2–6, we assumed no direct effect of partner smoking on participation, and

https://doi.org/10.1017/S2040174424000199 Published online by Cambridge University Press

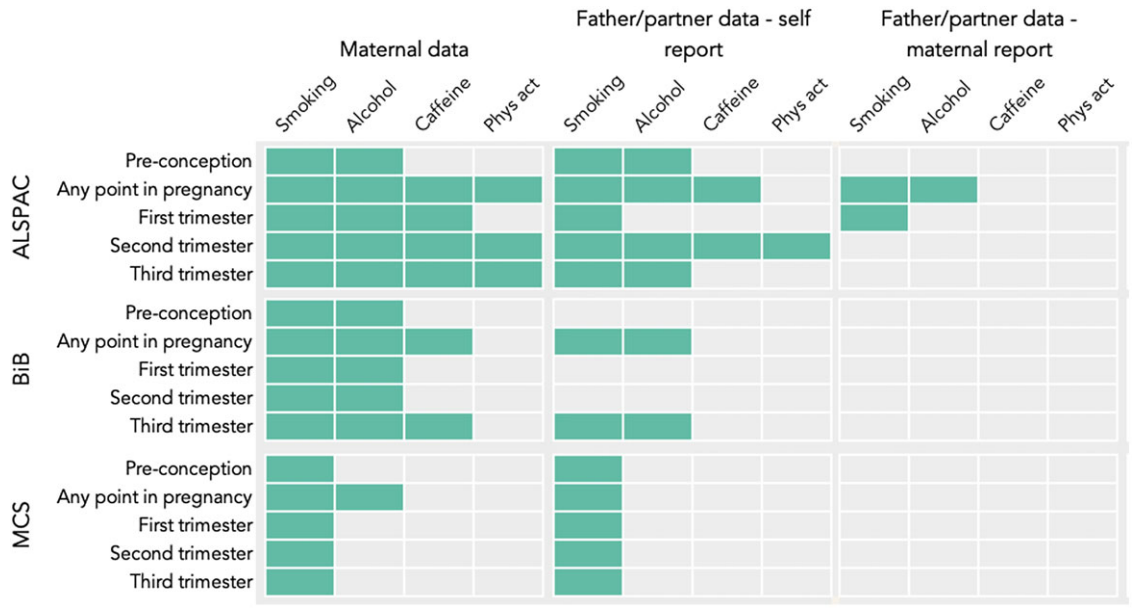


Figure 3. Availability of data on parental prenatal health behaviours in Avon Longitudinal Study of Parents and Children, Born in Bradford and Millennium Cohort Study.

set participation effects for SEP and BMI to be such that the odds ratio for participation per one category increase in SEP (coded as a categorical variable with four categories), and the odds ratio for participation per one standard deviation increase in BMI were equal to 1.5, 2, 5, 10 and 20, respectively. Some values are large, example, in Scenario 1, the odds ratio for participation per unit increase in partner SEP as reported by mothers was 1.58. Nevertheless, we chose to use these values to explore the impact of selection bias for a range of different strengths of association between partner traits and participation. In scenarios 7–11, we also assumed an effect of partner smoking on participation, with an odds ratio of 0.67, 0.5, 0.2, 0.1 and 0.05 respectively (see Supplementary Methods). Again, these values were selected to represent different magnitudes of association between partner smoking and participation (we used odds ratios lower than 1 since smoking is negatively correlated with participation in ALSPAC).

In additional simulations, reported in the Supplementary Material, we also explored the impact of outcome-participation confounding and smoking-general health interactions on selection bias.

Estimands: The associations of maternal and partner smoking with offspring birthweight, adjusted for each other and for both parents' SEP but not for general health.. We did not adjust for parents' general health, as a perfect measure of parents' general health is unlikely to be available in real datasets, therefore applied analyses would not be able to fully adjust for that variable.

Methods: Estimates were obtained from a linear regression model, fitted only using data on families with participating partners, with offspring birthweight as the outcome and maternal smoking, partner smoking, maternal SEP and partner SEP as covariates.

Performance measures: Absolute and relative bias, model-based standard errors, empirical coverage of 95% confidence intervals, and bias-eliminated coverage. Results were averaged across 1000 repetitions of the simulation.

We refer to the Supplementary Methods for more details about our simulation design and its ADEMP formulation.

Applied analysis of potential selection bias

To assess the extent of selection bias in real data, we conducted linear regression analyses and compared the effect estimates of maternal smoking on birthweight in ALSPAC, BiB and MCS using the total sample and samples stratified by partner participation status, both with and without adjustment for partner smoking.

Analyses were performed in R version 4.0.3 or Stata version 15, and code for our simulation study is available at [https://github.com/agkatzionis/Parental-smoking-and-birthweight.git].

Results

Partner participation

Of total pregnancies in each cohort there were the following with a cohort participating partner: In ALSPAC, of 14,472 pregnancies, 12,997 (78%); In MCS, of 18,241 pregnancies, 13,145 (71%); In BiB, of 11,538, 3,131 (27%).

Available data on parental prenatal health behaviours

In all three cohorts, data on partner prenatal health behaviours were less detailed and collected less frequently than data on maternal prenatal health behaviours (Figure 3).

Family characteristics by partner participation

Figure 4 shows the pattern of association between family characteristics and partner participation. Consistently across all cohorts (where data available), mothers were more likely to have a participating partner if the woman was white, had a degree, or was cohabiting with her partner before the birth of the child. Families where babies were born preterm or with a low birthweight were less likely to have a participating partner. Maternal report of partner measures showed that partners were more likely to participate if they were the biological parent, were white, had consumed alcohol during pregnancy, and had a degree. Those who were smokers or had no qualifications were less likely to participate. Other family characteristics, such as maternal alcohol consumption during

https://doi.org/10.1017/S2040174424000199 Published online by Cambridge University Press

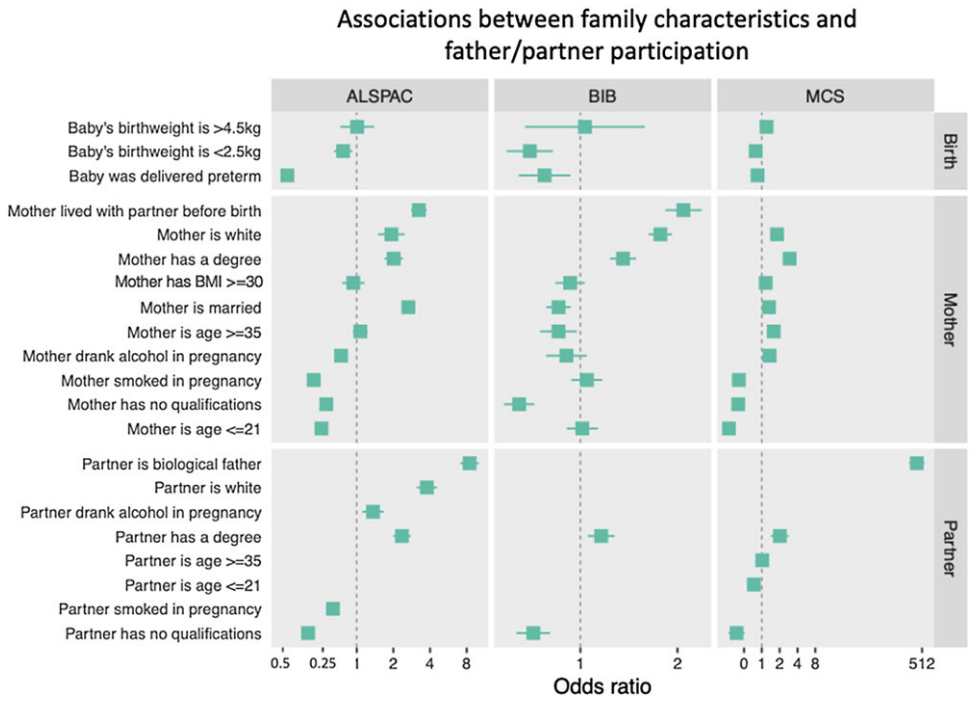


Figure 4. Association between family characteristics and participation of the partner in the cohort study. Squares represent the odds ratio of participation for each binary characteristic; horizontal lines show the 95% confidence intervals.

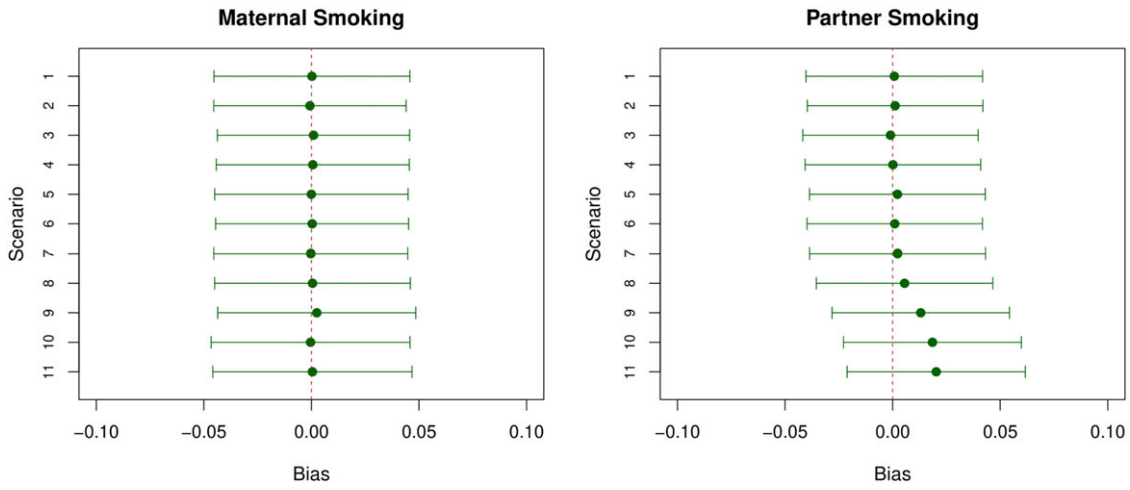


Figure 5. Forest plot summarising the results of our simulation study based on the Avon Longitudinal Study of Parents and Children cohort. Bias in estimated associations of maternal (left) and partner (right) smoking with (standardised) offspring birthweight. Estimates are obtained from a joint analysis of both maternal and partner smoking, and are also adjusted for both parents' socio-economic position. The true associations were -0.325 for maternal and -0.161 for partner smoking. The results are provided in numerical form in Supplementary Table 2.

pregnancy and parental age, had less consistent associations with partner participation across the three cohorts (Figure 4).

In all cohorts, higher SEP and being white were associated with higher rates of partner participation. Whereas higher SEP was associated with lower rates of smoking, being white was associated with higher rates of smoking (Figure 4), (97% white in ALSPAC; 75% white in MCS; 40% white in BIB). A large proportion of participants in BiB (51%) identify their ethnicity as South Asian. Rates of smoking in this group are lower than in the white group, particularly rates of maternal smoking during pregnancy (3% in South Asian mothers, 34% in white European mothers; 30% in South Asian partners, 34% in white European partners).

Simulation study to explore the magnitude and direction of bias due to sample attrition

Figure 5 and Supplementary Table 2 show results of the simulation study to assess the extent to which missing partner data could introduce selection bias. For the associations of maternal smoking with offspring birthweight there was little evidence of selection bias in any of the eleven scenarios, with estimates being close to the true value of the maternal effect on smoking in our simulations (relative bias less than 1%). For partner associations, there was some evidence of selection bias in Scenarios 9–11, where all three partner traits (SEP, smoking and health) had strong effects on participation. The presence of bias can be justified theoretically, according

https://doi.org/10.1017/S2040174424000199 Published online by Cambridge University Press

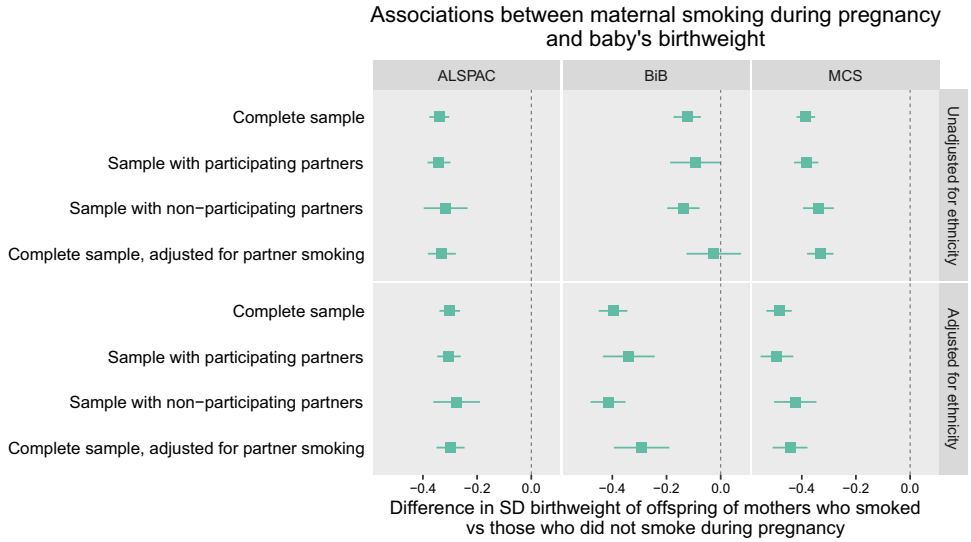


Figure 6. Effect estimates of maternal smoking during pregnancy on offspring birthweight (*z*-scores) in Avon Longitudinal Study of Parents and Children, Born in Bradford and Millennium Cohort Study. Linear regression was run on either the complete sample, samples stratified by partner participation, or the sample with participating partners adjusted for self-reported partner smoking sample (the latter model restricts the sample to participating partners). Models were run unadjusted (top row) and adjusted (bottom row) for maternal ethnicity.

to the theory of causal diagrams: using only data on families with participating partners induces bias through the path 'Partner smoking → Partner participation ← Partner Health → Birthweight'. In our simulations, the bias acted in a positive direction. The direction of bias can also be justified theoretically by noting that partner smoking leads to reduced participation while better partner health increases participation; therefore, among participating families, the association between partner smoking and general health will be overestimated[26] and so will the association of partner smoking with offspring birthweight. However, our results show that the magnitude of bias was small: across our 11 scenarios, the worst-case absolute bias was 0.02 and the worst-case relative bias was 12.6%. Similarly, the empirical coverage of 95% confidence intervals attained nominal levels in most simulation scenarios; in the worst-case scenario (partner smoking, Scenario 11) the empirical coverage was 83.6%. The bias would be eliminated if it was possible to adjust the analysis for partner general health, as we demonstrate in Supplementary Table 3.

There was little evidence of selection bias for either maternal or partner associations in the scenario where partner effects on selection were set according to the observed associations with maternal traits in ALSPAC (absolute bias 0.001, relative bias 0.5% for paternal/partner estimates).

Similar results were reported in additional simulations with outcome-participation confounding and smoking-general health interactions (Supplementary Tables 4 and 5), although the bias was stronger in these simulations.

Influence of missing partner data on partner-adjusted estimates of maternal effects using real (non-simulated) data

Figure 6 shows estimates of association between maternal smoking and offspring birthweight in ALSPAC, BiB and MCS (adjusted for partner smoking, and with and without adjustment for maternal ethnicity). We saw relatively stable effect estimates within cohorts regardless of whether we used the sample with or without participating partners, or adjusted or not for partner smoking. The exception was in BiB, where the association attenuated towards the null after adjustment for partner smoking.

Discussion

Available data on parental prenatal health behaviours

We have described the availability of data on partner prenatal health behaviours in three UK cohorts and explored how missing partner data might influence findings from studies of prenatal parental effects on child health. All four health behaviours (smoking, alcohol, caffeine, physical activity) were collected at more time points for mothers compared to partners. All cohorts included some self-report data on smoking during pregnancy for both mothers and partners. Physical activity had the least amount of data and was only available in one cohort (ALSPAC), at three timepoints for mothers and only one timepoint for partners.

Family characteristics by partner participation

Partners who participate in cohort studies have been shown to have higher SEP compared to partners who do not participate (see[27] for a review). This underrepresentation of certain demographics may contribute to selection bias. We tested whether participating partners (only available through the mother's participation) were systematically different from non-participating partners, using self and maternal report of partner characteristics (where we did not have self-reported data on non-participating partners), and also indirectly using maternal and birth characteristics, which are likely to be correlated with partner characteristics due to assortative mating, shared environments, and shared father-offspring genetics. Across all three cohorts we studied, pregnancies with participating partners were more likely to have a degree, be married and over 21 years old. This suggests partners who did not participate in these cohorts were likely to have a lower SEP than those who did. These findings are in agreement with a previous study[28] showing that working class fathers, less educated fathers and fathers from minority ethnic groups were underrepresented in the National Institute of Child Health and Human Development Study of Early Child Care. Moreover, babies of participating fathers were healthier with higher birthweights than babies of fathers who did not participate.[28] These findings are important for future researchers including fathers/partners within their study design, to be aware of the impact of and appropriately adjust for demographic factors.

https://doi.org/10.1017/S2040174424000199 Published online by Cambridge University Press

Influence of missing partner data in our real-data analyses, estimates of the effects of maternal smoking on offspring birthweight did not differ substantially when using the complete (unselected) sample, the sample where partners participated, or the sample where they did not. This was the case in all three cohorts suggesting that, although adjusting for missing partner data can introduce selection bias, in the examples we investigated, the biasing effect appears to be small. This was confirmed by a simulation study. The importance of mutually adjusting maternal and partner exposures in negative control analyses has been demonstrated[29,30] and it is possible that lower recruitment of partners may then result in bias of both maternal (real) and partner (negative control) results; in our simulations, however, only the latter was observed.

There were relatively stable effect estimates within cohorts if we used the sample with or without paternal participation or adjusted for paternal smoking, except for within BiB where the association attenuated towards the null after adjustment for paternal smoking, this is likely due to confounding by ethnicity. In all cohorts, being white and a higher SEP were associated with higher partner participation, however, higher SEP was associated with lower rates of smoking, and being white was associated with higher rates of smoking. Because BiB is ethnically diverse, the association between maternal smoking and birthweight may be confounded by ethnicity and the results are therefore presented both unadjusted and adjusted by ethnicity. This confounding relationship has little effect on the estimates generated in ALSPAC and MCS because participants in these cohorts are mostly white (97% white in ALSPAC, 75% white in MCS). However, only 40% of BiB are white and a large proportion of participants (51%) identify their ethnicity as South Asian. Rates of smoking in this group are lower than in the white group, particularly rates of maternal smoking during pregnancy (3% in South Asian mothers, 34% in white European mothers; 30% in South Asian partners, 34% in white European partners). Therefore, ethnicity (or rather, cultural factors proxied by ethnicity) is a major source of variation in maternal smoking data in BiB.

Strengths and limitations

The strengths of our study include the use of three large UK cohorts that are commonly used to study maternal, and increasingly, partner effects on offspring health.[15,31] The use of simulations alongside real-data analysis allowed us to explore effects of selection bias. The simulation parameters were, where possible, informed by ALSPAC data to be more realistic. Our study has a number of limitations. We focused on three cohorts with distinct partner recruitment and follow-up strategies, as examples. Other cohorts may have different partner recruitment strategies and the impact of this on selection bias may differ from the ones we have studied. We also included only UK cohorts but cohorts in other countries may have different challenges (e.g., different cultures of research participation) which may have different impacts (e.g., on the extent and direction of selection bias). The simulation study looked at an example research question, estimating the effect of maternal and partner smoking on offspring birthweight, but selection bias for different exposures and outcomes are likely to have different magnitudes and direction. Within the simulation study we used BMI as a proxy for general health, but the relationship between BMI and other variables in our analysis may not be the same as these variables' relationship with general health. For example, while we would expect SEP and general health to be positively correlated, the SEP–BMI association that we observed in ALSPAC was negative (because individuals with a high SEP are less likely to have very high BMI). The aim of this paper however was to compare maternal-offspring associations with the equivalent partner-offspring associations. We also only explored a prespecified research question and a small set of scenarios (utilising the health behaviour, smoking, as most readily studied across cohorts), future research may seek to investigate using additional measures. Lastly, within non-participating partners, we were unable to distinguish between partners who declined to participate and partners of mothers who stated that they did not have a partner. There may be differences in characteristics between these groups which could impact any bias.

Conclusions

We agree with others[6,28] that more studies of partner effects (in their own right) and studies of maternal effects that adjust for partner data, are needed. However, as with all epidemiological studies, care should be taken designing and interpreting results from studies using partner data. Based on our findings, our recommendations for future data collection within new cohorts and studies including of the effects of prenatal parental health behaviours are as follows. Firstly, efforts should be made to maximise partner recruitment where possible. Secondly, align data collection for partners and mothers (depth/breadth). Third, collect maternal-reported partner measures where partner self-report is unavailable, particularly given its proven accuracy.[32] Fourth, gather paternity data as biological and non-biological partners impact offspring health through different mechanisms. Fifth, data should be collected concerning paternity, because biological and non-biological partners affect offspring health through different mechanisms. Lastly, in studies on prenatal exposures we suggest adjusting maternal effects for partner data (and vice versa), using mutual adjustment to prevent bias from assortative mating.[30]

Our study has highlighted discrepancies in the breadth and depth of cohort study data on the health behaviours of partners compared to mothers. Theoretically, missing father/partner data could introduce selection bias in studies of parental effects on child health. However, using real and simulated data, we have shown that any biasing effect of partner selection to birth cohort studies is likely to be small, although this may differ across cohort studies in different populations and with different selection mechanisms. We hope that our findings and recommendations encourage more research that collects and analyses data on partners.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S2040174424000199.

Acknowledgements. ALSPAC: We are extremely grateful to all the families who took part in this 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. MCS: We are grateful to the Centre for Longitudinal Studies (CLS), UCL Social Research Institute, for the use of these data and to the UK Data Service for making them available. However, neither CLS nor the UK Data Service bear any responsibility for the analysis or interpretation of these data.

Financial support. This work was supported by the UK Medical Research Council and the Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide which provide core support for ALSPAC. This publication is the work of the authors and GCS will serve as guarantor for the contents of this

paper. This work was performed in the UK Medical Research Council Integrative Epidemiology Unit, jointly funded by the University of Bristol and the UK MRC. KEE and GCS were financially supported by an MRC New Investigator Research Grant awarded to GCS (grant code MR/S009310/1). GCS is also financially supported by the European Joint Programming Initiative 'A Healthy Diet for a Healthy Life' (JPI HDHL, NutriPROGRAM project, UK MRC MR/S036520/1). AG, KT (MC UU 00,032/02) and DAL (MC UU 00,032/05) are supported by the Medical Research Council Integrative Epidemiology Unit at the University of Bristol.

Competing interests. None.

Ethical standard. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national guidelines on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008, and has been approved by the institutional committees: the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

References

1. Sharp GC, Lawlor DA, Richardson SS. It's the mother!: how assumptions about the causal primacy of maternal effects influence research on the developmental origins of health and disease. *Soc Sci Med*. 2018; 213, 20–27. DOI: 10.1016/j.socscimed.2018.07.035.
2. Sharp GC, Schellhas L, Richardson SS, Lawlor DA. Time to cut the cord: recognizing and addressing the imbalance of DOHaD research towards the study of maternal pregnancy exposures. *J Dev Orig Health Dis*. 2019; 10, 509–512. DOI: 10.1017/S2040174419000072.
3. Easey KE, Wootton RE, Sallis HM, *et al*. Characterization of alcohol polygenic risk scores in the context of mental health outcomes: within-individual and intergenerational analyses in the avon longitudinal study of parents and children. *Drug Alcohol Depend*. 2021; 221, 108654. DOI: 10.1016/j.drugalcdep.2021.108654.
4. Schellhas L, Haan E, Easey KE, *et al*. Maternal and child genetic liability for smoking and caffeine consumption and child mental health: an intergenerational genetic risk score analysis in the ALSPAC cohort. *Addiction*. 2021; 116(11), 3153–3166. DOI: 10.1111/add.15521.
5. Camlin NJ, Sobinoff AP, Sutherland JM, *et al*. Maternal smoke exposure impairs the long-term fertility of female offspring in a murine model. *Biol Reprod*. 2016; 94(2), 39. DOI: 10.1095/biolreprod.115.135848.
6. Braun JM, Messerlian C, Hauser R. Fathers matter: why it's time to consider the impact of paternal environmental exposures on children's health. *Curr Epidemiol Rep*. 2017; 4(1), 46–55. DOI: 10.1007/s40471-017-0098-8.
7. Sharp GC, Lawlor DA. Paternal impact on the life course development of obesity and type 2 diabetes in the offspring. *Diabetologia*. 2019; 62(10), 1802–1810. DOI: 10.1007/s00125-019-4919-9
8. Easey KE, Sharp GC. The impact of paternal alcohol, tobacco, caffeine use and physical activity on offspring mental health: a systematic review and meta-analysis. *Reprod Health*. 2021; 18(1), 214. DOI: 10.1186/s12978-021-01266-w.
9. Abbasi J. The paternal epigenome makes its mark. *JAMA*. 2017; 317(20), 2049–2051, 10.1001/jama.2017.1566.
10. Rando OJ. Daddy issues: paternal effects on phenotype. *Cell*. 2012; 151(4), 702–708. DOI: 10.1016/j.cell.2012.10.020.
11. Day J, Savani S, Krempley BD, Nguyen M, Kitlinska JB. Influence of paternal preconception exposures on their offspring: through epigenetics to phenotype. *Am J Stem Cells*. 2016; 5(1), 11–18.
12. Davey Smith G. Assessing intrauterine influences on offspring health outcomes: can epidemiological studies yield robust findings? *Basic Clin Pharmacol Toxicol*. 2008; 102, 245–256.
13. Tyrrell J, Huikari V, Christie JT, *et al*. Genetic variation in the 15q25 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) interacts with maternal self-reported smoking status during pregnancy to influence birth weight. *Hum Mol Genet*. 2012; 21(24), 5344–5358. DOI: 10.1093/hmg/dds372.
14. Taylor AE, Carslake D, de Mola CL, *et al*. Maternal smoking in pregnancy and offspring depression: a cross cohort and negative control study. *Sci Rep*. 2017; 7(1), 2579. DOI: 10.1038/s41598-017-11836-3.
15. Easey KE, Timpson NJ, Munafò MR. Association of prenatal alcohol exposure and offspring depression: a negative control analysis of maternal and partner consumption. *Alcohol Clin Exp Res*. 2020; 44(5), 1132–1140. DOI: 10.1111/acer.14324
16. Sanderson E, Macdonald-Wallis C, Davey Smith G. Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration. *Int J Epidemiol*. 2018; 47(2), 587–596. DOI: 10.1093/ije/dyx213.
17. Tolonen H, Ahonen S, Jentoft S, Kuulasmaa K, Heldal J, European Health Examination Pilot Project. Differences in participation rates and lessons learned about recruitment of participants–the European health examination survey pilot project. *Scand J Public Health*. 2015; 43(2), 212–219. DOI: 10.1177/1403494814565692.
18. Dunn KM, Jordan K, Lacey RJ, Shapley M, Jinks C. Patterns of consent in epidemiologic research: evidence from over 25,000 responders. *Am J Epidemiol*. 2004; 159(11), 1087–1094. DOI: 10.1093/aje/kwh141
19. 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(1), 97–110. DOI: 10.1093/ije/dys066.
20. 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. DOI: 10.1093/ije/dys064.
21. 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 Res*. 2019; 4, 51. DOI: 10.12688/wellcomeopenres.15132.1.
22. Wright J, Small N, Raynor P, *et al*. Cohort profile: the born in Bradford multi-ethnic family cohort study. *Int J Epidemiol*. 2013; 42(4), 978–991. DOI: 10.1093/ije/dys112.
23. NICE. *Diabetes in pregnancy: management from preconception to the postnatal period*, <https://www.nice.org.uk/guidance/ng3/chapter/Recommendations#gestational-diabetes> 2020).
24. Connelly R, Platt L. Cohort profile: UK millennium cohort study (MCS). *Int J Epidemiol*. 2014; 43(6), 1719–1725. DOI: 10.1093/ije/dyu001.
25. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019; 38(11), 2074–2102. DOI: 10.1002/sim.8086
26. Jiang Z, Ding P. The directions of selection bias. *Stat Probabil Lett*. 2017; 125, 104–109.
27. Doyle O, Weller BE, Daniel SS, Mayfield A, Goldston DB. Overcoming barriers to fathers' participation in clinically relevant research: recommendations from the field. *Soc Work Res*. 2016; 40, 260–264.
28. Soubry A. Epigenetics as a driver of developmental origins of health and disease: did we forget the fathers? *Bioessays*. 2018; 40(1). DOI: 10.1002/bies.201700113.
29. Robinson MR, Kleinman A, Graff M, *et al*. Genetic evidence of assortative mating in humans. *Nat Hum Behav*. 2017; 1, 0016. DOI: 10.1038/s41562-016-0016
30. Madley-Dowd P, Rai D, Zammit S, Heron J. Simulations and directed acyclic graphs explained why assortative mating biases the prenatal negative control design. *J Clin Epidemiol*. 2020; 118, 9–17. DOI: 10.1016/j.jclinepi.2019.10.008.
31. Cohen JM, Wood ME, Hernández-Díaz S, Ystrom E, Nordeng H. Paternal antidepressant use as a negative control for maternal use: assessing familial confounding on gestational length and anxiety traits in offspring. *Int J Epidemiol*. 2019; 48(5), 1665–1672. DOI: 10.1093/ije/dyz170
32. Passaro KT, Noss J, Savitz DA, Little RE. Agreement between self and partner reports of paternal drinking and smoking. The ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood.. *Int J Epidemiol*. 1997; 26, 315–320. DOI: 10.1093/ije/26.2.315.