Received: 8 April 2021 | Revised: 7 September 2021 | Accepted: 8 September 2021

DOI: 10.1111/ppe.12821

METHODOLOGY ARTICLE

Paediatric and Perinatal Epidemiology WILEY

Lost to follow-up in the Norwegian mother, father and child cohort study

Kristine Vejrup[1,2] | Per Magnus[1] | Maria Magnus[1]

[1]Norwegian Institute of Public Health, Oslo, Norway

[2]University of Agder, Kristiansand, Norway

Correspondence
Kristine Vejrup, Center for Fertility and Health, Norwegian Institute of Public Health, P.O. Box 222, Skoyen, NO-0123 Oslo, Norway.
Email: Kristine.vejrup@fhi.no

Abstract

Background: The aim of pregnancy cohorts was to understand causes and development of health and disease throughout the life course. A major challenge in cohort studies is to avoid selection bias from loss to follow-up.

Objective: The aim of this study was to describe what characterises drop out from the Norwegian Mother, Father and Child Cohort Study (MoBa), and provide a resource to inform the interpretation of results from analysis of cohort data.

Methods: We estimated loss to follow-up in subsets of participants that responded to questionnaire waves in MoBa through an eight-year period and described characteristics of participants who responded to follow-ups. Within each wave of questionnaires, we estimated two exposure-outcome associations: the relationship between maternal smoking during pregnancy and offspring birthweight, and between educational level and pre-pregnancy body mass index (BMI). We explored the use of inverse probability weighting to correct the bias due to loss to follow-up.

Results: Participants who continued to respond were older, higher educated, less likely to smoke and had lower BMI. We observed a decline in participation of current smokers from 22.3% to 17.5%, and participants who reported an unplanned pregnancy dropped from 19.2% to 16.4%. There was a gradual decline in the inverse relationship between maternal smoking during pregnancy and offspring birthweight with increasing follow-up information, indicating that selection bias due to drop out resulted in lower effect estimates. For the relationship between parental educational level and BMI, the inverse association increased with amount of follow-up information, indicating that the selection bias resulted in higher effect estimates. Inverse probability weighting did not completely correct the estimates for bias due to loss to follow-up.

Conclusions: Participants who remain cohort members are different from subjects who drop out. Users of large cohorts should be aware of selective loss to follow-up and consider imputation or weighting to account for loss to follow-up when analysing questionnaire responses.

KEYWORDS
birth cohort, father and child cohort study, loss to follow-up participation, MoBa, The Norwegian mother

1 | BACKGROUND

There is evidence that the environment during the prenatal and early postnatal period may be critically important for the development of chronic disorders in adulthood. Pregnancy cohorts provide one of the best methodologies for studying aetiology and disease mechanisms throughout the life course.[1,2] With the advance of large-scale biobanking and lower costs of genotyping, population-based pregnancy cohorts have expanded their research potential even further.[3]

All prospective cohort studies are at risk of participant drop out, and when loss to follow-up of many participants occurs, the internal validity of the study may be affected. Systematic differences related to the outcome or risk factors of interests between those who drop out and those who remain in the study may introduce bias in associations of interest.[4]

Studies of loss to follow-up from two large contemporary European pregnancy cohorts, the Danish National Birth cohort (DNBC) and the Avon Longitudinal Study of Parents and Children (ALSPAC), show that loss to follow-up was non-random.[5,6] The presence and magnitude of bias depend on the exposure-outcome relationship under examination.[5] Selective loss to follow-up does not necessarily bias all associations.[7]

The Norwegian Mother, Father and Child Cohort Study (MoBa) is one of the largest pregnancy cohorts in the world. It is a nationwide population-based cohort that collects data to provide insight into how exposures throughout the life course may influence health and disease development. Pregnant women and their partners were invited to participate, and the cohort now contains information from pregnancy, birth, childhood and adolescence for more than 100,000 Norwegian children and their families.

If there is a systematic loss to follow-up in MoBa, it may lead to errors in conclusions drawn from research using this cohort. Our aim was to describe what characterises the drop out and provide a resource to inform the interpretation of results from analysis of data from the cohort. We estimated how loss to follow-up influenced two well-known relationships measured at baseline: the difference in birthweight according to maternal smoking during pregnancy and the socio-economic gradient in body mass index.

Synopsis

Study question

The aim was to describe what characterises the drop out in the Norwegian Mother, Father and Child Cohort Study, and provide a resource to inform interpretation of results from the cohort.

What's already known

Participants who remain cohort members are different from subjects who drop out, and a major challenge in cohort studies is to avoid selection bias from loss to follow-up.

What this study adds

We found that participants who continued to respond to questionnaires were older, higher educated, less likely to smoke and had lower BMI. Researchers who use data from the cohort study should be aware of the selective loss to follow-up that might influence their research questions when analysing data from questionnaire responses and consider imputation or weighting to account for the loss to follow-up.

2 | METHODS

2.1 | Cohort selection; The Norwegian mother, father and child cohort study (MoBa)

MoBa recruited pregnant women and their partners around 18th gestational weeks across Norway between 1999 and 2008.[8] The overall participation rate was 41%. All participants gave a written informed consent. Since parents could participate with more than one pregnancy, the cohort includes 95,000 mothers, 75,000 fathers and 114,500 children. These comprise of approximately 76,000 women who contributed with one pregnancy, 14,000 women who contributed with two pregnancies, and 2500 women who contributed with three or more pregnancies. Collection of data continued through pregnancy and later at intervals following birth, mainly by self-reported questionnaires, but with more extensive data collection in a few sub-studies. Information on the child's birth record is available from the Medical Birth Registry of Norway (MBRN) through linkage using unique personal identifiers. The study population consist of those participants who are registered as active in the cohort and have received a questionnaire at a given timepoint from pregnancy to the child is 8 years.

2.2 | Exposure

Relevant background characteristics were examined in relation to loss to follow-up.

In the baseline questionnaire answered by pregnant women at recruitment in gestational week 18, Q1, women reported their own and their partner´s educational level (less than high school, high school, up to four years of college and more than four years of college), income level (<200,000 NOK, 200,000–400,000 NOK and more than 400,000 NOK), height (continuous), weight (continuous) and smoking status (never, former and current), in addition to their use of folate supplements (yes, no) and whether the pregnancy was planned (yes, no). We studied self-reported information on the mother´s chronic diseases including asthma, hay fever, eczema, insulin-dependent diabetes, non-insulin-dependent diabetes, rheumatoid arthritis, psoriasis, Crohn's disease/ulcerative colitis, epilepsy, migraine, hypertension and high cholesterol (yes, no). From the MBRN record, we used information on maternal and paternal

age at delivery, sex of the child (male, female), birthweight (continuous in grams), gestational age (continuous in weeks estimated by ultrasound or last menstrual period for those with missing ultrasound measures), Apgar score at one minute (continuous), Apgar score at five minutes (continuous), pre-eclampsia (yes, no) and gestational diabetes (yes, no). We defined preterm birth as gestational age less than 37 completed weeks. Small-for-gestational age was defined as a birthweight below the 10th percentile according to offspring sex and gestational week of delivery. Low Apgar scores at one and five minutes were defined as below seven (yes, no). Baseline characteristics for participants from all questionnaires are available in Tables S1–S3.

2.3 | Outcome

We estimated the proportions of participants who had completed follow-up questionnaires at each wave of data collection. We show simple descriptive statistics of background characteristics among individuals who responded to follow-up questionnaire, to describe how the underlying composition of the population changes over time due to loss to follow-up.

2.4 | Statistical analysis

To illustrate the effect of selective follow-up on association measures, we used two well-known associations: the relationship between maternal smoking during pregnancy and birthweight[9,10] and socio-economic status and body mass index (BMI).[11,12] We estimated these two associations, which were based on information collected at baseline or at birth, within the restricted samples of participants who had responded to the later questionnaires. We show the association with responding to the 8-year questionnaire only among participants who were sent the questionnaire. Illustrative DAGs of the associations between maternal smoking and birthweight, and the association between education and BMI are included in the supplement (Figure S1). We explored the use of inverse probability weighting to correct the estimates for the potential bias due to loss to follow-up.[13] The weights were estimated using all baseline characteristics presented in Table 2. All analyses were conducted using Stata version 15 (Statacorp, Texas).

2.5 | Missing data

Missing data at each wave of data collection reflect both deaths of participants and unwillingness to continue to respond to questionnaires. The development of follow-up questionnaires in the cohort study has not always been corresponding with the ageing of the cohort children. This has resulted in a fraction of the participants not being sent some questionnaires due to the child being past the appropriate age of the respective data collection, thus leading to missing data. This pertains particularly to the 5-year questionnaire, as this questionnaire was not developed before a proportion of children had already turned 5 years. Missing data due to loss to follow-up are described in Table 1.

2.6 | Ethics approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Regional Ethics Committee of South East Norway (11.03.2019/ 2019/411).

3 | RESULTS

The response rates for each questionnaire are shown in Table 1. The rates remained high (85% or higher) for the first five questionnaires administered up until the child was six months of age. After this, the rate

TABLE 1 Response rate for completed follow-up questionnaires

	Time of completion of questionnaire	Participants who received questionnaire	Participants who returned questionnaire	Response rate %
Q1	18 gestational week[a]	112,580	102,174	90.8
Q2	22 gestational week[a]	107,507	97,251	90.5
QF	15 gestational week -Father[a]	8810	78,325	90.2
Q3	30 gestational week[a]	103,555	94,227	91.0
Q4	Child 6 months	105,815	89,752	84.8
Q5	Child 18 months	105,395	76,450	72.5
Q6	Child 3 years	100,325	58,876	58.7
Q5y	Child 5 years	77,776	41,636	53.5
Q7y	Child 7 years	101,196	54,825	54.2
Q8y	Child 8 years	93,464	43,649	46.7
QF2	All fathers 2015	77,321	49,485	64.0

[a]The unit is the pregnancy. For other questionnaires, the unit of the analysis is the child.

TABLE 2 Distribution of baseline parental characteristics at baseline and participants still in the cohort after 8 years, mean differences and risk ratios for responding to the 8-year questionnaire according to these characteristics

	Baseline questionnaire *N* = 102,174	Answered 8-year questionnaire *N* = 43,649	Mean differences and risk ratios for responding to the 8-year questionnaire
Baseline parental characteristics	**Mean (SD)**	**Mean (SD)**	**Mean difference (95%CI)**
Maternal age	30.2 (4.6)	30.7 (4.4)	0.98 (0.79, 1.17)
Paternal age	32.7 (5.4)	33.1 (5.3)	0.67 (0.34, 1.00)
Maternal BMI	23.9 (4.1)	23.8 (3.9)	−0.34 (−0.39, −0.29)
Paternal BMI	25.8 (3.3)	25.7 (3.2)	−0.21 (−0.26, −0.17)
	N (%)	**N (%)**	**RR (95%CI)**
Maternal educational level			
Less than high school	8218 (8.0)	2084 (4.8)	1.00 (Reference)
High school	30031 (29.4)	10707 (24.5)	1.33 (1.28, 1.38)
Up to 4 years of higher education	40754 (39.9)	19077 (43.7)	1.69 (1.62, 1.75)
> 4 years of higher education	22620 (22.1)	11184 (25.6)	1.75 (1.68, 1.82)
Missing	551 (0.5)	597 (1.37)	n.a.
Paternal educational level			
Less than high school	10626 (10.4)	3607 (8.26)	1.00 (Reference)
High school	39927 (39.1)	15764 (36.1)	1.13 (1.10, 1.16)
Up to 4 years of higher education	26305 (25.8)	12164 (27.9)	1.30 (1.26, 1.33)
More than 4 years of higher education	21842 (21.4)	10494 (24)	1.34 (1.30, 1.38)
Missing	3474 (3.4)	1620 (3.71)	n.a.
Marital status			
Married/cohabitated	97545 (95.5)	42168 (96.6)	1.00 (Reference)
Other	4187 (4.1)	1404 (3.2)	0.81 (0.77, 0.84)
Missing	442 (0.4)	77 (0.2)	n.a.
Maternal parity			
0	45638 (44.7)	19,889 (45.6)	1.00 (Reference)
1	36190 (35.4)	15,404 (35.3)	0.99 (0.97, 1.00)
2	15502 (15.2)	6556 (15.0)	1.00 (0.98, 1.02)
3 or higher	4402 (4.3)	1723 (3.9)	0.95 (0.91, 0.98)
Missing	442 (0.4)	77 (0.2)	n.a.
Maternal income			
<200,000	29,369 (28.7)	10,291 (23.6)	1.00 (Reference)
200,000– 400.000	57,679 (56.5)	26,256 (60.2)	0.84 (0.82, 0.85)
>400,000	11,444 (11.2)	5453 (12.5)	1.01 (0.99, 1.03)
Missing	3682 (3.6)	1649 (3.8)	n.a.
Paternal income			
<200,000	10,638 (10.4)	3896 (8.9)	1.00 (Reference)
200,000– 400,000	52,683 (51.6)	22,108 (50.6)	1.10 (1.07, 1.12)
>400,000	31,552 (30.9)	14,742 (33.8)	1.15 (1.12, 1.18)
Missing	7301 (7.1)	2903 (6.7)	n.a.
Mother or father, not native Norwegian speakers			
No	86,641 (84.8)	38,202 (87.5)	1.00 (Reference)
Yes	11,135 (10.9)	4347 (10.0)	0.94 (0.92,0.96)
Missing	4398 (4.3)	1100 (2.5)	n.a.

(Continues)

TABLE 2 (Continued)

	N (%)	N (%)	RR (95%CI)
Maternal smoking status			
Never	51,715 (50.6)	23,420 (53.7)	1.00 (Reference)
Former	27,124 (26.5)	11,960 (27.4)	0.93 (0.91, 0.94)
Current	22,757 (22.3)	7646 (17.5)	0.77 (0.75, 0.78)
Missing	578 (0.6)	623 (1.4)	n.a.
Paternal smoking			
Never/former	80,595 (78.9)	35,496 (81.3)	1.00 (Reference)
Current	20761 (20.3)	7425 (17.0)	0.86 (0.84,0.88)
Missing	818 (0.8)	728 (1.7)	n.a.
Maternal use of folate supplements during pregnancy			
No	20,991 (20.5)	6569 (15.0)	1.00 (Reference)
Yes	81,160 (79.4)	36,645 (84.0)	1.24 (1.21, 1.26)
Missing	23 (0.1)	435 (1.0)	n.a.
Whether the pregnancy was planned or not			
Yes	81,245 (79.5)	35,585 (81.5)	1.00 (Reference)
No	19664 (19.2)	7162 (16.4)	0.87 (0.85, 0.88)
Missing	1265 (1.2)	902 (2.1)	n.a.

has gradually decreased and was down to around 50% for the last three questionnaires sent out when the child was five, seven and eight years of age. Notably, the 5-year questionnaire was initiated after a proportion (20%) of the children were too old to contribute to this data collection.

Table 2 shows parental characteristics at baseline including age, socio-economic factors and life-style characteristics. Retention in the cohort was higher for participants who were older, lived with a partner, had higher educational level, had higher income level, were non-smokers, had planned pregnancies and more often had used folic-acid supplements during pregnancy.

The proportion of participants, whose pregnancy was not planned, dropped from 19.2% at baseline to 16.4% at the 8-year follow-up. At baseline, 4.1% of the participants did not live with a partner. This proportion of single parents at baseline that still participated at the 8-year follow-up decreased to 3.2%. We also observed a decrease in the participation of mothers who reported smoking during pregnancy, from 22.3% at baseline to 17.5% after eight years. There was no strong evidence of a selection due to having one or more chronic diseases (Table S3).

There was no change in the distribution of offspring sex between the population available at baseline and those with follow-up information at 8 years (Table 3). However, there was a slightly reduced proportion of children born preterm and children with APGAR scores at one minute less than seven, and at five minutes less than seven (Table 3). The reduction in birthweight associated with maternal smoking in pregnancy varied from a decrease of 169 grams among respondents to the baseline questionnaire to a decrease of 114 grams among respondents to the 5-year questionnaire (Table 4).

The loss to follow-up was higher among participants with a lower educational level (Table 2). The association between maternal education and pre-pregnancy BMI was stronger for respondents to the later questionnaires (Table 5). A similar pattern was found for the relationship between paternal education and BMI (Table 6). Inverse probability weighting was performed for all analysis but did not completely correct the differences in estimates due to loss to follow-up.

4 | COMMENT

4.1 | Principal findings

In this study, we found that after eight years of follow-up to a large cohort study, the distributions of a series of characteristics for the remaining respondents differed from the original distributions at baseline. This selective follow-up appeared to lead to a lower degree of association between maternal smoking and offspring birthweight, and a higher association between BMI and educational level, when only the remaining sample was analysed. These analyses do show systematic association differences with increasing loss to follow-up over time, but at a modest magnitude, and the general relationships of the associations were maintained. We attempted to use inverse probability weighting to correct for the bias due to loss to follow-up for our illustrative examples. However, for our examples, this approach did not completely correct the estimates. It is possible that loss to follow-up in the cohort results in a more homogenous study population, perhaps resulting in less confounding. The results do not necessarily imply that one of the associations is more correct than the other.

4.2 | Strengths of the study

Our findings are consistent with other studies on loss to follow-up in pregnancy cohorts.[14-16] The participants who are most

TABLE 3 Distribution of pregnancy outcomes according to participation response to 8-year follow-up, and risk ratios (RRs) for responding to the 8-year questionnaire, according to maternal disease and birth outcomes

	Baseline questionnaire N=104014	8-year questionnaire N=43649	RR for answering the 8-year questionnaire
Pregnancy outcome	**N (%)**	**N (%)**	**RR (95% CI)**
Preeclampsia			
No	99,377 (95.5)	41,913 (96.0)	1.00 (Reference)
Yes	4195 (4.0)	1659 (3.8)	0.95 (0.92, 0.99)
Missing	442 (0.4)	77 (0.2)	n.a.
Gestational diabetes			
No	102,684 (98.7)	43,254 (99.1)	1.00 (Reference)
Yes	888 (0.9)	318 (0.7)	0.85 (0.78, 0.93)
Missing	442 (0.4)	77 (0.2)	n.a.
Sex			
Male	52,990 (50.9)	22,238 (50.9)	1.00 (Reference)
Female	50,379 (48.5)	21,331 (48.9)	1.01 (0.99, 1.02)
Missing	645 (0.6)	80 (0.2)	n.a.
Preterm birth			
No	96,109 (92.4)	40,824 (93.5)	1.00 (Reference)
Yes	7022 (6.8)	2569 (5.9)	0.95 (0.92, 0.98)
Missing	883 (0.8)	256 (0.6)	n.a.
Small for gestational age			
No	93,240 (89.6)	39,410 (90.3)	1.00 (Reference)
Yes	10,055 (9.7)	4141 (9.5)	0.98 (0.96, 1.00)
Missing	719 (0.7)	98 (0.2)	n.a.
APGAR at 1 minute <7			
No	97,214 (93.5)	41,234 (94.5)	1.00 (Reference)
Yes	5902 (5.7)	2288 (5.2)	1.00 (0.97, 1.03)
Missing	898 (0.9)	127 (0.3)	n.a.
APGAR at 5 minutes <7			
No	101,527 (97.6)	43,061 (98.7)	1.00 (Reference)
Yes	1597 (1.5)	462 (1.1)	0.95 (0.90, 1.02)
Missing	890 (0.9)	126 (0.3)	n.a.

likely to not respond to follow-up questionnaires and drop out of the cohort have characteristics that are linked to lower socioeconomic status. These participants are known to be more difficult to engage in longitudinal studies.[17,18] Answering long and frequent questionnaires are time consuming and can make the participants lose interest in the cohort study if they do not feel that they contribute to something important. In our opinion, it is therefore important to keep questionnaires short and concise. The technological development using smartphones has also opened up for easier and less resource-demanding contact with participants and also allows participants to respond to questionnaires when it is convenient for them (while on the bus, waiting in line for their coffee, etc.).

The Danish birth cohort is the one that most closely resembles MoBa. Of mothers that were invited to participate in the 7-year follow-up, 60% responded. These women were slightly older, were more likely to have a normal BMI, were more likely to be non-smokers and were more likely to have planned pregnancies compared with non-respondents.[5] In the ALSPAC study, the response rate was 48% 12 years after recruitment.[6] The drop out from the cohort was related to being a single parent, low education, financial difficulties and being raised in a large family where the mother smoked.[7] The Western Australian Pregnancy Cohort (RAINE) has been ongoing for two decades and reports a participation rate of 70% at the 5-year assessment with a decline to a 42% participation at the 20-year assessment. There were greater attrition among socially disadvantaged participants with a fall out of younger mothers who were not married at recruitment.[19] In the Generation R Study, the participation rate when the children were five years was 85% and decreased to 76% at age 13. The mothers who still participated in the study at follow-up were older, higher educated and more frequently of Dutch nationality.[20]

TABLE 4 Reduction in birthweight (grams) for offspring of mothers who smoked occasionally or daily during pregnancy, according to response to questionnaires

	Occasional smoking	Weighted analysis Occasional smoking	Daily smoking	Weighted analysis daily smoking
Questionnaire	Mean difference in grams (95% CI)	Mean difference in grams (95% CI)	Mean difference in grams (95% CI)	Mean difference in grams (95% CI)
Q1	−30 (−59, −2)	n.a.	−169 (−184, −153)	n.a.
Q2	−32 (−61, −3)	−52 (−86, −19)	−160 (−175, −144)	−141 (−160, −133)
Q Father	−44 (−79, −10)	−62 (−102, −23)	−162 (−180, −143)	−155 (−177, −133)
Q3	−26 (−55, 2)	−39 (−71, −7)	−154 (−169, −138)	−137 (−155, −119)
Q4	−16 (−45, 14)	−35 (−68, −2)	−144 (−160, −128)	−130 (−149, −111)
Q5	−19 (−52, 15)	−30 (−67, 8)	−141 (−159, −123)	−123 (−145, −102)
Q6	−45 (−84, −6)	−57 (−100, −14)	−147 (−168, −126)	−134 (−159, −109)
Q5y	−30 (−80, 21)	−29 (−84, 27)	−114 (−143, −86)	−112 (−145, −79)
Q7y	−33 (−73, 7)	−49 (−96, 2)	−140 (−164, −117)	−139 (−168, −110)
Q8y	−42 (−88, 3)	−58 (−110, 7)	−144 (−171, −117)	−136 (−169, −103)
Q2 Father	−57 (−111, −2)	−80 (−144, −15)	−126 (−158, −94)	−142 (−181, −102)

4.3 | Limitations of the data

A major challenge when starting a prospective cohort is to include a representative selection of the population that you want to study to avoid selection bias. In MoBa, 41% of the invited pregnant women consented to participation, and even though the study participants were found to not be representative of the entire Norwegian population, a previous study examining the role of selection bias into the cohort found that associations between exposures of interest and pregnancy outcomes did not significantly differ from the associations estimated among all deliveries in Norway during the MoBa recruitment period.[21]

The next big challenge in prospective cohorts is to keep as many of the participants in the study for as long as possible. Bias due to loss to follow-up can compromise the internal validity of exposure-outcome associations and are often driven by an overrepresentation of participants with high education level and socio-economic status.[22]

Research on participation in cohort studies shows that there are several socio-demographic and individual factors that indicate whether a participant is prone to stay in a study and keep responding to questionnaires.[23] When examining participants still in the MoBa cohort eight years after recruitment, we saw a trend towards an overrepresentation of older parents, and an underrepresentation of parents with lower educational attainment, lower income level, smokers and unplanned pregnancies. However, presence of chronic diseases during pregnancy and unfavourable birth outcomes were less predictive of further participation.

4.4 | Interpretation

Our observations suggest the traits that are related to the participants' socio-economic status may be affected by selection. Specifically, socio-economic differences seem to be higher in our cohort than in the general population. The likelihood and severity of any selection bias should be evaluated on a case-by-case basis to understand the extent to which participant loss to follow-up influences the internal validity of associations of interest as opposed to only influencing the external validity/generalisation of results.[24-26]

This study provides an overview of the changes in the distribution of important background characteristics among participants who are retained in the MoBa cohort. We acknowledge that the modest selection bias illustrated in the two examples shown here, the relationship between maternal smoking and offspring birthweight and the relationship between educational level and BMI, do not necessary indicate that all associations are impacted by selection bias. The aim was to show the differences in established associations according to the amount of follow-up information available in the MoBa cohort and to illustrate the potential for selection bias. While selection bias may not affect all research questions of interest in the cohort, it is important that the researchers planning to use data from the cohort are carefully considering this issue when planning their analysis and that they consider imputation or weighting to account for the loss to follow-up. This information can be used to guide researchers using this pregnancy cohort in their evaluation of how selection might have influenced their associations of interest and prompt any relevant sensitivity analysis. We hope that more researchers will consider using both multiple imputation and inverse probability weighting to account for the selection present in the cohort due to loss to follow-up.[27] By using inverse probability weights or multiple imputation, we are shifting the assumption of missing completely at random to missing at random. We are still requiring that we meet the assumption of missing at random, meaning that we have enough information available to inform the weight or imputation.[28] Notably, the exact variables which should be included in the generation of weights for inverse probability weighting or in the imputation model will vary

TABLE 5 Reduction in pre-pregnancy body mass index (BMI) for mothers with high school or college education compared to mothers with less than high school education

	High school education	Weighted analysis High school education	Up to 4 years of college	Weighted analysis Up to 4 years of college	>4 years of college	Weighted analysis >4 years of college
Questionnaire	Mean difference in BMI (95% CI)	Mean difference in BMI (95% CI)	Mean difference in BMI (95% CI)	Mean difference in BMI (95% CI)	Mean difference in BMI (95% CI)	Mean difference in BMI (95% CI)
Q1	−0.13 (−0.23, −0.03)	n.a.	−0.80 (−0.90, −0.70)	n.a.	−1.52 (−1.62, −1.42)	n.a.
Q2	−0.15 (−0.25, −0.04)	−0.17 (−0.27, 0.06)	−0.83 (−0.93, −0.73)	−0.87 (−1.01, −0.73)	−1.54 (−1.65, −1.44)	−1.61 (−1.74, −1.42)
Q Father	−0.13 (−0.25 −0.01)	−0.10 (−0.32 −0.03)	−0.85 (−0.96, −0.73)	−0.83 (−0.99, −0.67)	−1.60 (−1.72, −1.48)	−1.58 (−1.75, −1.47)
Q3	−0.12 (−0.23, −0.01)	−0.13 (−0.28, 0.02)	−0.82 (−0.92, −0.72)	−0.85 (−0.99, −0.70)	−1.53 (−1.64, −1.42)	−1.58 (−1.72, −1.43)
Q4	−0.17 (−0.28, −0.05)	−0.16 (−0.32, 0.00)	−0.85 (−0.96, −0.74)	−0.87 (−1.02, −0.72)	−1.56 (−1.68, −1.45)	−1.61 (−1.77, −1.45)
Q5	−0.25 (−0.37, −0.12)	−0.27 (−0.45, −0.09)	−0.93 (−1.05, −0.81)	−0.99 (−1.16, −0.71)	−1.63 (−1.75, −1.50)	−1.71 (−1.89, −1.54)
Q6	−0.30 (−0.45, −0.15)	−0.33 (−0.54, −0.11)	−1.00 (−1.15, −0.85)	−1.05 (−1.26, −0.85)	−1.71 (−1.86, −1.55)	−1.78 (−1.89, −1.57)
Q5y	−0.17 (−0.36, −0.03)	−0.21 (−0.49, 0.07)	−0.87 (−1.05, −0.68)	−0.91 (−1.19, −0.64)	−1.55 (−1.75, −1.36)	−1.62 (−1.89, −1.35)
Q7y	−0.23 (−0.39, −0.08)	−0.24 (−0.47, −0.02)	−0.91 (−1.06, −0.75)	−0.97 (−1.19, −0.75)	−1.59 (−1.74, −1.43)	−1.70 (−1.92, −1.48)
Q8y	−0.35 (−0.53, −0.16)	−0.44 (−0.71, −0.18)	−1.00 (−1.18, −0.82)	−1.13 (−1.39, −0.87)	−1.70 (−1.88, −1.51)	−1.87 (−2.13, −1.61)
Q2 Father	−0.56 (−0.77, −0.34)	−0.41 (−0.74, −0.09)	−1.27 (−1.48, −1.06)	−1.18 (−1.49, −0.86)	−1.97 (−2.19, −1.76)	−1.92 (−2.23, −1.60)

TABLE 6 Reduction in paternal baseline body mass index (BMI) according to level of education in accordance to response to follow-up questionnaires. The reference group consists of participants with less than high school education

	High school education	Weighted High school education	Up to 4 years of college	Weighted Up to 4 years of college	>4 years of college	Weighted >4 years of college
Questionnaire	Mean difference (95% CI)	Mean difference (95% CI)	Mean difference (95% CI)	Mean difference (95% CI)	Mean difference (95% CI)	Mean difference (95% CI)
Q1	−0.14 (−0.21, −0.07)	n.a.	−0.58 (−0.65, −0.51)	n.a.	−1.05 (−1.13, −0.98)	n.a.
Q2	−0.15 (−0.23, −0.08)	−0.19 (−0.29, −0.10)	−0.60 (−0.68, −0.52)	−0.62 (−0.72, −0.53)	−1.06 (−1.14, −0.98)	−1.11 (−1.20, −1.09)
Q Father	−0.23 (−0.32, −0.15)	−0.25 (−0.36, −0.14)	−0.70 (−0.79, −0.62)	−0.71 (−0.82, −0.60)	−1.20 (−1.29, −1.10)	−1.20 (−1.31, −1.09)
Q3	−0.15 (−0.23, −0.08)	−0.18 (−0.28, −0.09)	−0.60 (−0.68, −0.52)	−0.62 (−0.71, −0.52)	−1.08 (−1.16, −1.00)	−1.12 (−1.22, −1.02)
Q4	−0.17 (−0.24, −0.09)	−0.17 (−0.27, −0.07)	−0.61 (−0.69, −0.53)	−0.61 (−0.71, −0.51)	−1.11 (−1.19, −1.02)	−1.12 (−1.22, −1.02)
Q5	−0.21 (−0.29, −0.12)	−0.20 (−0.31, −0.08)	−0.66 (−0.75, −0.57)	−0.65 (−0.77, −0.54)	−1.14 (−1.23, −1.05)	−1.16 (−1.27, −1.04)
Q6	−0.23 (−0.33, −0.13)	−0.22 (−0.35, −0.09)	−0.69 (−0.79.−0.59)	−0.67 (−0.80.−0.54)	−1.22 (−1.32, −1.11)	−1.21 (−1.34, −1.09)
Q5y	−0.22 (−0.34, −0.09)	−0.20 (−0.36, −0.03)	−0.71 (−0.83, −0.58)	−0.70 (−0.87, −0.53)	−1.22 (−1.35, −1.09)	−1.23 (−1.39, −1.06)
Q7y	−0.25 (−0.36, −0.15)	−0.29 (−0.44, −0.15)	−0.68 (−0.79, −0.58)	−0.74 (−0.88, −0.59)	−1.18 (−1.29, −1.07)	−1.25 (−1.40, −1.11)
Q8y	−0.23 (−0.35, −0.12)	−0.24 (−0.40, −0.08)	−0.70 (−0.82, −0.58)	−0.70 (−0.86, −0.54)	−1.20 (−1.32, −1.07)	−1.23 (−1.39, −1.07)
Q2 Father	−0.41 (−0.56, −0.26)	−0.48 (−0.69, −0.27)	−0.96 (−1.11, −0.80)	−1.06 (−1.27, −0.85)	−1.43 (−1.58, −1.28)	−1.55 (−1.79, −1.34)

according to the association of interest.[29] We therefore encourage researchers to evaluate this for their specific research question.

One should also keep in mind that all participants that remain in the cohort can be linked to registries for health outcomes. This opens for studies where the same phenotype can be studied both among responders and among all cohort members and invites comparisons of exposure-outcome associations. Researchers using information from the national health registries in Norway to ascertain outcomes for MoBa participants would minimise their risk of bias due to loss to follow-up.

The results presented here highlight the importance of minimising loss to follow-up in pregnancy cohorts. This includes continuously providing participants with updated information about research discoveries made using the information they have provided to the study, so that they can see that their efforts have contributed important new insights. It is also important to give participants the opportunity to share their opinions of planned research projects, for example in participant focus group, and to be part of the shaping of new research ideas. This will help give them a sense of ownership of the cohort and hopefully prompt their continued participation. Unfortunately, such efforts are often undervalued by researchers and lack necessary funding.

5 | CONCLUSIONS

Loss to follow-up is a problem for all prospective cohort studies. It is important to bear in mind how this may have introduced selection bias for the particular relationship of interest. Researchers that use data from the cohort study should consider imputation or weighting to account for the loss to follow-up.

ACKNOWLEDGEMENTS

We are grateful to all the participating families in Norway who take part in this ongoing cohort study. We would like to thank Elin Alsaker at the MoBa data unit for facilitation of the data file used in the analysis. We also want to acknowledge all the researchers and administrative employees at the Norwegian Institute of Public Health that has contributed to making MoBa a successful cohort study.

AUTHOR CONTRIBUTIONS

Maria Magnus has performed all analysis. All three authors has contributed in discussions and interpretations of the analysis. Kristine Vejrup has been in charge of the design of the article, and all three authors has active contributed to the contents.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from NIPH. Restrictions apply to the availability of these data, which were used under license for this study.

ORCID

Kristine Vejrup https://orcid.org/0000-0002-7448-9529
Maria Magnus https://orcid.org/0000-0002-0568-3774

REFERENCES

1. Magnus P, Haug K, Nystad W, Skjaerven R. The Mother and Child Cohort Study will give new answers. *Tidsskr nor Laegeforen*. 2006;126(13):1747-1749.
2. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med*. 2008;5(3):e67.
3. Stoltenberg C. Large cohorts: toward routine databases for public health science. *Am J Public Health*. 2016;106(9):1536-1537.
4. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: How much is too much? *Eur J Epidemiol*. 2004;19:751-760.
5. Greene N, Greenland S, Olsen J, Nohr EA. Estimating bias from loss to follow-up in the Danish national birth cohort. *Epidemiology (Cambridge, Mass)*. 2011;22(6):815-822.
6. 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(1):111-127.
7. Wolke D, Waylen A, Samara M, et al. Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *Br J Psychiatry*. 2009;195(3):249-256.
8. Magnus P, Birke C, Vejrup K, et al. Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol*. 2016;45(2):382-388.
9. Rayfield S, Plugge E. Systematic review and meta-analysis of the association between maternal smoking in pregnancy and childhood overweight and obesity. *J Epidemiol Community Health*. 2017;71(2):162-173.
10. Witt SH, Frank J, Gilles M, et al. Impact on birth weight of maternal smoking throughout pregnancy mediated by DNA methylation. *BMC Genom*. 2018;19(1):290.
11. Vieira LS, Bierhals IO, Vaz JDS, Meller FO, Wehrmeister FC, Assunção MCF. Socioeconomic status throughout life and body mass index: a systematic review and meta-analysis. *Cad Saude Publica*. 2019;35(10):e00125518.
12. Grabner M. BMI trends, socioeconomic status, and the choice of dataset. *Obesity Facts*. 2012;5(1):112-126.
13. Sheikh K. Investigation of selection bias using inverse probability weighting. *Eur J Epidemiol*. 2007;22(5):349-350.
14. Charles MA, Thierry X, Lanoe J-L, et al. Cohort Profile: The French National cohort of children ELFE: birth to 5 years. *Int J Epidemiol*. 2019.
15. Heude B, Forhan A, Slama R, et al. Cohort Profile: The EDEN mother-child cohort on the prenatal and early postnatal determinants of child health and development. *Int J Epidemiol*. 2015;45(2):353-363.
16. Guxens M, Ballester F, Espada M, et al. Cohort Profile: The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project. *Int J Epidemiol*. 2011;41(4):930-940.
17. Wadsworth ME, Mann SL, Rodgers B, Kuh DJ, Hilder WS, Yusuf EJ. Loss and representativeness in a 43 year follow up of a national birth cohort. *J Epidemiol Community Health*. 1992;46(3):300-304.
18. Krishnan S, Wu K, Smurzynski M, et al. Incidence rate of and factors associated with loss to follow-up in a longitudinal cohort of antiretroviral-treated HIV-infected persons: an AIDS Clinical Trials Group (ACTG) Longitudinal Linked Randomized Trials (ALLRT) analysis. *HIV Clin Trials*. 2011;12(4):190-200.
19. White SW, Eastwood PR, Straker LM, et al. The Raine study had no evidence of significant perinatal selection bias after two decades of follow up: a longitudinal pregnancy cohort study. *BMC Pregnancy and Childbirth*. 2017;17(1):207.
20. Kooijman MN, Kruithof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. *Eur J Epidemiol*. 2016;31(12):1243-1264.
21. Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr Perinat Epidemiol*. 2009;23(6):597-608.

22. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol*. 2007;17(9):643-653.
23. Mostafa T, Wiggins R. The impact of attrition and non-response in birth cohort studies: a need to incorporate missingness strategies. *Longitudinal and Life Course Studies*. 2015;6(2):16.
24. Ebrahim S, Davey SG. Commentary: Should we always deliberately be non-representative? *Int J Epidemiol*. 2013;42(4):1022-1026.
25. Richiardi L, Pizzi C, Pearce N. Commentary: Representativeness is usually not necessary and often should be avoided. *Int J Epidemiol*. 2013;42(4):1018-1022.
26. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol*. 2013;42(4):1012-1014.
27. Perkins NJ, Cole SR, Harel O, et al. Principled approaches to missing data in epidemiologic studies. *Am J Epidemiol*. 2018; 187(3):568-575.
28. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol*. 2019;48(4):1294-1304.
29. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2011;22(3):278-295.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Vejrup K, Magnus P, Magnus M. Lost to follow-up in the Norwegian mother, father and child cohort study. *Paediatr Perinat Epidemiol*. 2021;00:1–10. https://doi.org/10.1111/ppe.12821