Clinical Therapeutics

Analyzing missing data in perinatal pharmacoepidemiology research: methodological considerations to limit the risk of bias

--Manuscript Draft--

Manuscript Number:	CLINTHER-D-19-00196R1
Article Type:	Commentary
Section/Category:	Pharmacoepidemiology
Keywords:	Missing data; pharmacoepidemiology; pregnancy; medication safety
Corresponding Author:	Angela Lupattelli, PhD NORWAY
First Author:	Angela Lupattelli, PhD
Order of Authors:	Angela Lupattelli, PhD
	Mollie Wood, PhD
	Hedvig Nordeng, PhD
Abstract:	Pharmacoepidemiological studies on the safety of medication during pregnancy are all susceptible to missing data, i.e., data that should have been recorded but for some reason were not. Missing data are ubiquitous, irrespective of the data source utilized. Bias can arise when incomplete information on confounders, outcome measures, pregnancy duration, or even cohort selection criteria, are used to estimate prenatal exposure effects that would be obtained from the fully observed data, if these were available for each mother-child dyad. This commentary describes general missing data mechanisms and methods, and illustrates how missing data were handled in recent medication in pregnancy research, according to the utilized data source. We further present one applied example on missing data analysis within the Norwegian Mother and Child Cohort Study (MoBa), and finally illustrate how the causal diagram framework can be of aid in assessing risk of bias due to missing data in perinatal pharmacoepidemiology research. We recommend applied researchers to limit missing data during data collection, to carefully diagnose missingness, to apply strategies for missing data mitigation under different assumptions, and finally to include evaluations of robustness results under these assumptions. Following this set of recommendations can aid future perinatal pharmacoepidemiology research in avoiding the problems that result from failure to consider this important source of bias.

UiO : University of Oslo
Faculty of Mathematics and Natural Sciences/School of pharmacy

Clinical Therapeutics

Editor Dr. Cynthia Willey-Temkin

Date: 14 October 2019

Submission of manuscript

Dear Editor,

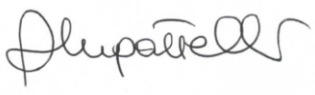
We would like to submit our revised manuscript entitled "***Analyzing missing data in perinatal pharmacoepidemiology research: methodological considerations to limit the risk of bias***" as a commentary in *Clinical Therapeutics*. This is a contribution for the special issue in pharmacoepidemiology, "Willey-Temkin Specialty Update".

In this revised version of the commentary, we have tried to address all comments raised by the Reviewers and the Editor, and amended the revised manuscript accordingly. As advised by the Reviewers and also by the Editors, we have now simplified the Introduction and Figure 1, and made the text clearer to enhance clarity.

No parts of the work have been published earlier or are under consideration for publication elsewhere. The authors have no financial or other relationships that might lead to a conflict of interest. The submitted version of the commentary has been read and approved by all authors.

We look forward to hearing from you at your earliest convenience.

On behalf of all authors,



Angela Lupattelli, PhD
Department of Pharmacy, University of Oslo,
PO Box 1068 Blindern, 0316 Oslo, Norway
E-mail: angela. lupattelli@farmasi.uio.no;
phone: +47 41343628 ; fax : +47 22854402

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

We thank the Editors and the Reviewers for the opportunity to revise our manuscript and for the valuable feedback provided. We have tried to address and implement in the revised manuscript all comments rose by the Reviewers.

To facilitate readability, we have split some of the comments, and numbered them within each Reviewer. Our replies to each individual comment are provided below and numbered accordingly.

EDITOR'S SPECIFIC COMMENTS:

This paper addresses a key methodologic issue in pharmacoepidemiology, and has the potential to provide guidance for numerous, future epidemiologic studies.

The reviewers express the need for more background information on the MoBa study, and the addition of more extensive, summary guidance for epidemiologists in the conclusions.
Reviewer 1 has made suggestions regarding grammar, and a software based grammar check could be used to address the need for specific changes.

Both reviewers have expressed the need for more conclusive statements about the results presented in Table 2 and Figure 1. I agree with reviewer 3 that Figure 1 may be too complex, as many readers will be learning about the DAG framework for the first time. Simplifying Figure 1, and perhaps adding summary text for each panel, could improve the use of this figure as a teaching tool.

The advice given in the section on "implications for applied researchers" reflects the authors' extensive experience with missing data approaches, and both reviewers feel that this section could be expanded to further guide epidemiologists in choosing the best approach.

Reply: Dear Editor, we have tried to improve the manuscript by addressing all these points. Specific replies to the Reviewers are provided below. To summarize, we have simplified Figure 1 and removed multiple parts of the Introduction, since these seemed to have caused confusion in one of the Reviewers. We hope the simplified form of the Introduction is now easier to understand for the readership of the journal *Clinical Therapeutics.*

Reviewer #1:

Comment 1: Major flaw: poor English especially in the introduction. I lost my interest after reading just the introduction.

The language is not only grammatically incorrect but also scientifically weak.

Reply 1: We respectfully disagree with the reviewer that the language used in our submitted article is scientifically weak. The language and terminology used in the manuscript are based on epidemiological text books and peer reviewed scientific articles.[1-3] To improve clarity throughout the manuscript, we have now simplified multiple parts, in particular the Introduction. All corrections made are visible in the revised manuscript using track change mode.

Comment 2: The introduction does not justify or highlight the novelty of this study. It failed to explain the issue and reason we need this study.

Reply 2: In the original version Introduction we highlighted the fact that missing data are ubiquitous, often overlooked, and that missing data can introduce bias in epidemiological studies if not adequately handled. In the revised version of the manuscript, we have simplified this section so that it is easier for the readership of the journal *Clinical Therapeutics* to appraise why this commentary is important, and how this work can guide applied researchers in handling missing data problems. All corrections made are visible in the revised manuscript using track change mode.

Comment 3: Although grammatical errors exist throughout the article, it is a major concern for the introduction section. It appears 'introduction' and rest of the article were written by 2 distinct persons.

Reply 3: The article was written by the same author, and the text was edited multiple times by one of the co-authors (native English speaker with extensive research experience) to ensure this is grammatically correct. According to Reviewer#3, this is a well written review. The article has one been edited for grammatical correctness one more time, and all needed corrections have been made.

Comment 4: When writing about MoBa study, please provide few lines related to the data. Since data is the key ingredient here, few introductory lines about it would strengthen the study.

Reply 4: We have now provided more detail about the MoBa study, which reads as follows: “MoBa is a nation-wide, population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health, with recruitment occurring between 1999-2008.23 Pregnant women were recruited from all

over Norway at the time of their routine ultrasound at 17-18 weeks of gestation. Data were gathered prospectively by self-administered questionnaires. The cohort now includes 114500 children, 95200 mothers and 77300 fathers, all of whom are followed as long as they continue to participate in the study23 MoBa has a license from the Norwegian Data Inspectorate and approval from The Regional Committee for Medical Research Ethics. All participants gave their written informed consent prior to participation."

Comment 5: For this study author said, "the extent of missingness was not substantial between complete case study and approach I. Despite that complete case study would introduce bias." A proper justification is needed here to support the statement.

Reply 5: We have explained in the applied example that we conducted three sets of analyses: i) complete case analysis; ii) multiple imputation of missing data on the two SCL scales only (approach I); and iii) multiple imputation of missing data on the two SCL scales and on other maternal confounders (approach II). The difference between complete case analysis and approach I is lack of multiple imputation on confounders other than the SCL scales (i.e., maternal severity of depressive/anxiety symptoms in pregnancy). First, we first compared complete case analysis and approach I results. Second, because the missing data mechanism in our study seemed to be linked to maternal age and to the extent of completion of the SCL items, we speculate that failure to handle missing data on this specific measure (i.e., by conducting a complete case analysis), may have yield biases estimates.

To make this point clearer, we have rephrased as follows: "The extent of missing data on confounders other than the SCL between the complete-case and approach I (31.9% vs 24.1%) analysis was however not substantial. Hence, because in this example missing data seemed to relate to the extent of completion of the SCL items, we could not exclude the possibility that a complete case analysis approach would yield biased estimates."

Comment 6: For the advanced methodologies, author needs to provide the areas these advance methods are lacking. Author recommended some readings in this section. A brief conclusive statement would be much welcomed for this section along with a comparison across different advanced methods.

Reply 6: It is outside the scope of this commentary to review uptake of different advances methods across fields within epidemiology. Although the Reviewers suggestion is of relevance, our submitted manuscript is a Commentary and it has the specific focus on perinatal pharmacoepidemiology.

Comment 7: While specifying DAG, it would be better to explain why we need it despite advanced methods.

Reply 7: As stated in the manuscript and in prior literature,[4] DAGs can aid researchers in identifying biases arising from missing data, since these are not always obvious, and also support researchers in differentiating between mechanisms of missingness (e.g. MNAR vs MAR), which are not testable statistically. These points were already presented in the manuscript. To facilitate clarity and better understanding as to why DAGs are important, we have now re-written the section "DAG framework with missing data". The first sentence of this section explains why DAGs are important, as follows: "Directed acyclic graphs (DAGs) can provide helpful insights into potential biases from assuming various missingness mechanisms".

Comment 8: Multiple approaches/models explained by DAG were not tested anywhere. It seem all hypothetical. Is there any data support for the recommended models?

Reply 8: DAGs are based on subject knowledge, prior literature, and assumptions. There is no statistical test that informs researchers whether the assumed DAG is correct or not.[5] Only sensitivity analyses can support researchers within this context. The section entitled "DAG framework with missing data" explains very clearly what assumptions were made in each scenario.

Comment 9: IRB details needed.

Reply 9: There is no ethics approval requirement for this commentary in specific. The applied MoBa example stems from a published study, where authors can find information about ethics approval. When describing the MoBa study in the revised manuscript, information about ethics approval of MoBa in general, is now presented.

Comment 10: Please see specific comments in attached annotated document.

Reply 10: In the annotated document– Abstract section - the Reviewer suggested that the word "medication" was incorrect in the following sentence, and that it should be rephrased by "missingness": "The proposed set of recommendations can aid future medication in pregnancy research in avoiding the problems that result from failure to consider this important source of bias". We wish to inform that the

word “medication” is correct, since the paper addressed the topic of handling missing data in medication in pregnancy research. However, to avoid misunderstanding, we have now rephrased the wording “medication in pregnancy” with “perinatal pharmacoepidemiology” throughout the manuscript and abstract, and also changed the title of the manuscript accordingly.

In the annotated document, the following is stated in relation to page 1 of the Introduction, passage on current definitions of “missing data”:

“*The way of describing these different scenarios is absurd. It needs major revision. The discussion is based on these different scenarios. Unfortunately, author is failing to capture audience due to his/her inability to use correct words. Not even a single sentence is grammatically and scientifically correct. If you want to describe different scenarios, you can use a counter to distinguish one scenario from other. Use of words like ‘Indeed’, ‘Somewhat less broadly’, and ‘Finally’ should be done judicially. Words like ‘narrow definition’ and ‘bias’ have been thrown in the sentence without any context. There is a lack of clarity in each and every sentence*.”

We understand that standard terminology within causal inference and epidemiology may be far from the clinical terms, and have tried to the best of our knowledge to clarify and simplify the text. As stated earlier in replies no. 1 and 2, the entire Introduction has been rewritten and simplified. The original section describing missing data problems within the causal inference framework and types of biases has been removed.

In the annotated document, the Reviewer has the following comment on the Introduction: “*Author is jumping back-and-forth between multiple different aspect of the issue. First step should be to define missingness. How much missingness in the data is considered as missingness? At what point it starts becoming the concern? What are the possible concerns? Can they or have they been resolved/attempted to resolve in the past? What approaches are available? Are these approaches successful? Pros and cons in these approaches? And then you come to the discussion of further research*.”

As stated earlier in replies no. 1 and 2, the entire Introduction has been rewritten and simplified. The points raised in the latter comment above were presented in later sections of the manuscript, e.g. what are the possible mechanisms of missingness, what methods can be used to handle missing data, etc. We have also added citation to a newly published study[6] showing that the proportion of missing data should not guide the decision of whether to use or not multiple imputation methods, and added the following

in the section “Exploring extent and patterns of missingness”: “Recent research has also shown that the proportion of missing data should not be the major driver for the decision on how to handle missing data.17 In fact, even when the extent of missing data is large, results can still be unbiased provided that the MAR assumption is met and methods to handle missing data have been adequately applied”.

As highlighted in this latest work, and also discussed in the section “Implications for applied researchers”, there is no definite answer on which method is the best for all research context; rather, it is crucial to understand why data are missing, and based on these assumed mechanism of missing data, apply the most suitable method to handle missing data to reduce risk of biased results.

Reviewer #3: In general, this is a well written review. To make this article more meaningful to epidemiologists, I suggest the following revisions.

Reply: We thank the Reviewer for the positive feedback and valuable comments provided.

Comment 1: In general, it would be more helpful to further recommend how an epidemiologist can choose the best analytical approach for missing data. The authors did a good job in reviewing different approaches. However, it is still unclear to me what I should do next time when I face such an issue.

Reply 1: In the section "Implications for applied researchers" we listed multiple recommendations for researchers. Unfortunately, in the context of missing data handling, it is not possible to provide clear answers and recommendations that fit all research contexts. We have now tried to make this point clearer in the beginning of the section of the manuscript, by adding the following detail: "Based on our survey of the literature, we have several recommendations for applied researchers who need to analyze data with missing values. These recommendations are made bearing in mind that there is no missing data handling solution that fits all research contexts".

Comment 2: Table 2. It's good to see the results from different approaches. But, which one or ones are more appropriate for this particular example? How to decide which approach is more appropriate, certainly before trying different approaches and seeing the results?

Reply 2: Thank you for this comment. We have now tried to amend the wording of the section describing the applied example, so that our strategy to handle missing data becomes clearer to the reader. We first explored the extent of missing data and patterns of missingness by exposure and outcome strata; based on these descriptive results, we assumed that the underlying mechanism of missing data was related to maternal age and depressive/anxiety symptoms as measured by the SCL. Next, we carried out three sets of analyses that reflected what we had observed in the descriptive analysis of missing data on confounders.

Based on our assumptions that data were missing at random – in light of the descriptive patterns of missingness – and on prior literature, we concluded that the effect estimates obtained from approach II (multiple imputed data analysis) were less biased than those resulting from the complete case analysis. As mentioned in the article, it is not possible to test whether data are MAR, MCAR, or MNAR. Within this context, we also introduced DAGs, since these graphical tools can support researchers in differentiating between mechanisms of missingness.

We have added the following sentence as concluding remark for the motivating example section: “In the context of this motivating example, results from approach II were thereby considered as those least biased”.

Comment 3: Figure 1. Suggest revising Fig1 and related discussion to focus on the most reasonable scenario(s) for this particular question. Need to justify why one scenario or scenarios are very reasonable. Hope this exercise can teach a reader to decide which diagram is the best for his own question.

Reply 3: Thank you for this important comment. We have now revisited the section “DAG framework with missing data”, simplified Figure 1, and made the assumptions for each scenario presented in the Figure clearer. The mechanism of missing data on smoking status in pregnancy – as shown in the DAG examples - can happen under different circumstances, that are i) completely at random, ii) not at random, or iii) at random. Each scenario is reasonable, and depends by the specific study setting and context. We aimed to convey throughout the manuscript that there is no missing data handling solution (or assumptions about missing data mechanisms) that fits all research scenarios. The task of applied researchers is to first diagnose patterns of missingness in their specific study, and based on these, to make assumptions about missing data mechanism. These assumptions can be made more explicit with the aid of DAGs. After these steps, it is possible to apply strategies for missing data mitigation under different assumptions, and to evaluate the robustness of the results under these assumptions. We hope the revised part of the manuscript entitled “Implications for applied researchers” is now clearer for the readership.

References

1. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ.* 2009;338:b2393.
2. Rubin DB. Inference and missing data. *Biometrika.* 1976;63:581-92.
3. Hernan MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health.* 2004;58:265-71.
4. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. *Stat Methods Med Res.* 2012;21:243-56.
5. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10:37-48.
6. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* 2019;110:63-73.

Analyzing missing data in perinatal pharmacoepidemiology research: methodological considerations to limit the risk of bias

Lupattelli A,[1] Wood M,[2] Nordeng H[1,3]

[1]*PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, & PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Norway*
[2]*Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA*
[3]*Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway*

Corresponding author: Angela Lupattelli, PhD. PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, & PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Norway. PO Box 1068 Blindern, 0316 Oslo, Norway
e-mail: angela.lupattelli@farmasi.uio.no; phone: +47 41343628; fax: +4722854402

Contribution to authorship

HN, MW and AL have developed the research question, and drafted the commentary. All authors have all contributed to interpretation of the data presented in the commentary, and critically revised the commentary for intellectual content.

Conflict of Interest statement

The author HN declares no conflict of interest. The author MW declares no conflict of interest. The author AL declares no conflict of interest.

Introduction

Missing data is a global problem in human subjects research, and a serious threat to both validity and efficiency in effect estimation. The CONSORT 2010 Statement[1] advocates transparent reporting of the extent of missing data and how this issue was dealt with in the analysis, as this is crucial for readers to critically evaluate the study findings and potential biases. Recognition of the threat from these biases has resulted in calls for increased use of methods for dealing with missing data.[2] However, barriers exist that prevent applied pharmacoepidemiology researchers from assessing the potential gains to their own work, including understanding scenarios when simpler methods might be sufficient, or when complex approaches are needed. These barriers include a lack of resources that integrate missing data terminology and approaches with epidemiologic concepts, and a discussion of the strengths and weaknesses of the most common approaches.

We review the critical concepts for missing data problems, with the aim of integrating more traditional statistical language on missingness mechanisms with epidemiologic methods based on causal diagrams.[3] We have framed this commentary using examples from perinatal pharmacoepidemiology, including an applied example from the Norwegian Mother and Child Birth Cohort (MoBa): evaluating the effect of prenatal use of selective serotonin reuptake inhibitor (SSRI) antidepressants on preeclampsia in the presence of missing data on relevant confounders such as smoking status in gestation.

Missing data methods and mechanisms

Missing data are generally classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR)[2, 4, 5] as briefly described below.

Missing completely at random (MCAR)

Under this scenario, there are no systematic differences between the missing and the observed values.[2, 5] For example, if unexperienced health care personnel forget to ask about smoking during pregnancy, information about smoking will be missing at random in the pregnant woman's medical chart. The same occurs when study participants randomly forget to fill in or skip responses. There is no risk of bias with MCAR data, but there will be loss of precision.

Missing at random (MAR)

Missing at random is classified as any systematic difference between the missing values and the observed values, which can be explained by the observed data.[2, 5] For instance, depressed pregnant women may be less likely to report smoking than non-depressed.

Missing not at random (MNAR)

Missing not at random occurs in situations when systematic differences remain between the missing values and the observed values, even after the observed data are taken into account; missingness is thus related to unmeasured variables. For example, women who smoke during pregnancy may less likely report their smoking status. When missingness in a variable depends on the missing value itself, the unbiased estimate is not recoverable in observed data

Exploring extent and patterns of missingness

Although Little's test may help researchers to identify missingness that is MCAR vs. MAR, this test is not conclusive. In addition, no numerical diagnostics can differentiate MAR from MNAR. This means we are left with logical reasoning to inform us on the mechanism behind data missingness. Exploring the extent and pattern of missing data in one's own data sample (for

example by cross-tabulating variables with missing data against exposure and outcome), as well as using findings from previous studies and normative data (e.g. score distribution in a reference population) can give a hint of the underlying mechanism of missingness. This is important to appraise as it will guide decision making of missing data handling: the various approaches to missing data analysis require different assumptions about the underlying mechanisms.

Methods to handle missing data

Multiple methods for handling missing data are used in perinatal pharmacoepidemiology research. These methods fall into two broad categories: analyze the observed data (complete case analysis), or use some principled method for filling in the missing data (imputation). In complete case analysis (CCA), observations with missing data on relevant variables are dropped from the analysis. This approach will always produce unbiased results under the MCAR assumption, and may produce unbiased results under MAR or MNAR. CCA is commonly used in perinatal pharmacoepidemiology due to its simplicity (Table 1). In database linkage studies where study size is large, the loss of data has less impact on precision than in smaller size or different design studies.[6-10]

Single imputation comprises a set of techniques where missing value are replaced by a value from the observed data, for instance the mean or mode. The imputed values are assumed to be equal to the values that would have been observed if data had been complete. This method, however, underestimates uncertainty about the missing values and will therefore result in standard errors that are too small.[2, 5] In the study by Panchaud et al,[11] gestational age was conditionally imputed for 6% of the pregnancies based on the sample mean. In the study by Pasternak et al,[12] missing information on several baseline maternal characteristics was replace

using the mode. In longitudinal studies with repeated variable measurement, for example using questionnaires at several time points in pregnancy, the "last value carried forward" technique can be used to replace missing values with the last measured value of the individual, as done by Norby et al.[13] This method assumes that the observation of the individual remains the same since the last measured observation. Due to well-established shortcomings,[2, 5] single imputation techniques are less used in perinatal pharmacoepidemiology.

More advanced model-based methods for handling missing data have become more accessible to researchers in recent years through packages in standard statistical software. The two most common model-based methods are maximum likelihood using the expectation maximization (EM) algorithm and multiple imputation.[4, 14, 15] These are considered model-based methods since the researcher must make assumptions about the joint distribution of all variables in the model (including both outcomes and predictors).

Maximum likelihood methods using the EM algorithm uses each observation's available data to compute maximum likelihood estimates, rather than filling in the missing values. It runs until the algorithm converges to the "best fit" model for a set of data. The multiple imputation (MI) method fills in missing values by averaging from the distribution of the missing data given the observed data in a way that accounts for the uncertainty associated with the missing values. In MI by chained equations (MICE) a series of regression models are run whereby each variable with missing data is modeled conditional upon the other variables in the data.[14] At the end of one cycle, all missing values have been replaced with predicted values (imputations). The process is repeated for a number of cycles, with the imputations being updated at each cycle, finally resulting in one imputed dataset. The number of imputed datasets is generally between 5 and 20. Standard errors are calculated using Rubin's rules.[15, 16] The MI approach produces valid

estimates under the MAR assumption. This is a weaker assumption than MCAR and more likely to hold in observational studies. MI is a computationally intensive method which is increasingly used in perinatal pharmacoepidemiology research (Table 1).[7, 17-21] Yet, this needs to be applied after careful reflection about the missing data to avoid misleading conclusions. For a comprehensive review of multiple imputation for missing data in epidemiological studies we recommend the papers by Sterne et al,[5] and Perkins et al.[2] Recent research has also shown that the proportion of missing data should not be the major driver for the decision on how to handle missing data.[22] In fact, even when the extent of missing data is large, results can still be unbiased provided that the MAR assumption is met and methods to handle missing data have been adequately applied.

Missing data approaches in recent medication in pregnancy literature

Table 1 summarizes the reporting and handling of missing data in recent perinatal pharmacoepidemiology studies, by type of data source utilized. Of note, this study overview serves as common ground for appraising current methodological gaps, and it is not a comprehensive, systematic extract of the literature. Transparent reporting of the extent and handling of missing data, and the uptake of multiple imputation methods, remains limited. For instance, in multiple cases we computed the extent of missing data in a study using baseline characteristic data of the study sample based on numbers reported in each manuscript; in some studies, it was unclear what missing data approach was used. The majority of studies reported missing data on confounding variables, in different extent (from <1% to 65%) depending on the data source utilized.

On the basis of the missing data definition used by study authors, and the information reported, missing data do not seem to be a major problem in health registry, administrative claims, or pregnancy registry. This contrasts with studies set in birth cohorts, teratology information services, or general practice databases, which often have to contend with much higher levels of missingness, and with patterns that are likely to be informative. The substantial problem of missing data in these study types has promoted important methodological research on the topic,[23, 24] as well as a greater uptake of multiple imputation methods by researchers using this type of data (Table 1). Simpler approaches to handle missing data such as indicator variable, were not often reported in the papers we evaluated; this is encouraging given the well-established shortcomings of the method. Study authors rarely stated any assumptions they made about the underlying mechanism of missingness in the literature we reviewed.

DAG framework with missing data

Directed acyclic graphs (DAGs) can provide helpful insights into potential biases from assuming various missingness mechanisms. Figure 1 introduces a simplified causal model for the effect of prenatal SSRI exposure on preeclampsia. In this model, we assume a causal effect of depression severity on SSRI use and on smoking, and that smoking has an effect on preeclampsia risk. If these assumptions hold, we could estimate the effect of SSRI use on preeclampsia by conditioning on smoking and depression severity. If some fraction of the study sample lacks data on smoking, assumptions about the mechanism that explains the missingness will point to different strategies for analyzing our data. In Figure 1A, smoking is missing completely at random (MCAR), and we can fit a model for the effect of SSRI use on preeclampsia risk, adjusting for depression severity and smoking, in the complete case sample only, without risk of bias. Figures 1B shows that if missingness in smoking status is explained by depression severity,

we can also estimate unbiased effects in the complete case sample, as the covariates required for confounding control also block bias paths from missingness to the outcome. For missing data mechanisms where the missingness is predicted by the missing values, as in Figure 1C, or when the probability of being a complete case depends on the outcome, as in Figures 1D and 1E, complete case analysis will result in a biased estimate. Finally, the presence of an auxiliary variable (that is, a variable that predicts missingness but is unrelated to the causal mechanism being considered) allows for unbiased and efficient effect estimation via multiple imputation.

Applied example: prenatal antidepressant use and risk of preeclampsia

As a motivating example, we present recent work on the association between use of selective serotonin re-uptake inhibitor (SSRI) antidepressants during gestation and risk of late-onset preeclampsia, using data from the MoBa cohort study.[25] MoBa is a nation-wide, population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health, with recruitment occurring between 1999-2008.[23] Pregnant women were recruited from all over Norway at the time of their routine ultrasound at 17-18 weeks of gestation. Data were gathered prospectively by self-administered questionnaires. The cohort now includes 114500 children, 95200 mothers and 77300 fathers, all of whom are followed as long as they continue to participate in the study[23] MoBa has a license from the Norwegian Data Inspectorate and approval from The Regional Committee for Medical Research Ethics. All participants gave their written informed consent prior to participation.

In our study, we first explored patterns of missing data on important confounders by exposure and outcome strata. Missing values on these confounders ranged from 1-3% for maternal smoking and body mass index, to 7-8% for education and weight gain. Missing information on

maternal depressive and anxiety symptom severity in pregnancy, measured via the Hopkins Symptom Checklist-25 (SCL-25) at gestational week 17 (5 items, SCL-5) and 30 (8 items, SCL-8),[26, 27] was as follows: 5% and 10% on at least one of the SCL-5 or SCL-8 items, respectively; 15% total missing information simultaneously on either scale. However, only few women (< 3%) completed none or less than a half of the items composing the individual SCL scales. The missing data mechanism in our study seemed to be linked to maternal age and to the extent of completion of the SCL items, but importantly, it did not seem to be associated with the outcome, late-onset preeclampsia. Based on this and under the MAR assumption,[25] we conducted three sets of analyses: i) complete case analysis; ii) multiple imputation of missing data on the two SCL scales only (approach I); and iii) multiple imputation of missing data on the two SCL scales and on other maternal confounders (approach II). As shown in Table 2, the adjusted and weighted association measures were higher and less precise in the complete case analysis than in the other two sets. However, the results of the complete case analysis expanded to pregnancies with only SCL imputed values (approach I) were similar to those obtained in the fully imputed models (approach II). Increasing sample size and higher statistical power following multiple imputation can indeed explain these discrepancies. The extent of missing data on confounders other than the SCL between the complete-case and approach I (31.9% vs 24.1%) analysis was however not substantial. Hence, because in this example missing data seemed to relate to the extent of completion of the SCL items, we could not exclude the possibility that a complete case analysis approach would yield biased estimates.[5, 28, 29] In the context of this applied example, results from approach II were thereby considered as those least biased.

Implications for applied researchers

Methods for identifying, analyzing, and mitigating bias from missing data have advanced significantly in recent years, and are seeing greater uptake in applied perinatal pharmacoepidemiology research. Based on our survey of the literature, we have several recommendations for applied researchers who need to analyze data with missing values. These recommendations are made bearing in mind that there is no missing data handling solution that fits all research contexts.

First: where possible, limit missingness during collection of data. Recognize that no statistical method can make up for careful study design and data curation. Sometimes the assumptions a specific case of missing data require are simply so unrealistic that the effect estimate is unlikely to be informative. *Second*: carefully diagnose missingness, and use subject-area knowledge as well as exploratory and descriptive data analysis to understand plausible mechanisms of missingness. We suggest that a minimum standard for missing data analysis should be a complete reporting of missingness within strata of exposure and outcome. Researchers should consider the use of causal graphs to make their assumptions about missingness mechanisms explicit. *Third*: Be aware that the proportion of missing data should not be the major driver for the decision on how to handle missing data, but rather the assumed mechanism as to why data are missing. *Fourth*: include a statistical analyst with expertise in missing data methods. Inappropriate analyses using these complex methods can result in seriously biased results. *Finally*: apply strategies for missing data mitigation under different assumptions, and include evaluations of robustness results under these assumptions. For example, including both the complete case analysis and the multiply imputed results can allow readers to decide which estimate they prefer, depending on assumptions about the missingness mechanism.

Careful attention to missing data, and to the assumptions required for analysis of missing data, is necessary in all areas of research, including perinatal pharmacoepidemiology. With transparent reporting of the extent and assumed mechanisms of missing data, and by applying strategies for missing data mitigation under different assumptions, future research can avoid the problems that result from failure to consider this important source of bias.

Figure Legends

Figure 1. Causal diagrams showing relationships between prenatal SSRI use, maternal depression severity, preeclampsia, and smoking, as well as a binary indicator, $Miss_{SMK}$, denoting missing information in the smoking variable.

References

1. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340.
2. Perkins NJ, Cole SR, Harel O, Tchetgen Tchetgen EJ, Sun B, Mitchell EM, et al. Principled approaches to missing data in epidemiologic studies. *Am J Epidemiol.* 2018;187:568-75.
3. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. *Stat Methods Med Res.* 2012;21:243-56.
4. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *J R Stat Soc Series B Stat Methodol.* 1977;39:1-38.
5. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ.* 2009;338:b2393.
6. Bateman BT, Heide-Jorgensen U, Einarsdottir K, Engeland A, Furu K, Gissler M, et al. Beta-blocker use in pregnancy and the risk for congenital malformations: An international cohort study. *Ann Intern Med.* 2018;169:665-73.
7. Caniglia EC, Patel K, Huo Y, Williams PL, Kapetanovic S, Rich KC, et al. Atazanavir exposure in utero and neurodevelopment in infants: A comparative safety study. *AIDS.* 2016;30:1267-78.
8. Dhalwani NN, Szatkowski L, Coleman T, Fiaschi L, Tata LJ. Stillbirth among women prescribed nicotine replacement therapy in pregnancy: Analysis of a large uk pregnancy cohort. *Nicotine Tob Res.* 2019;21:409-15.
9. Elkjaer LS, Bech BH, Sun Y, Laursen TM, Christensen J. Association between prenatal valproate exposure and performance on standardized language and mathematics tests in school-aged children. *JAMA Neurol.* 2018;75:663-71.
10. Ernst A, Brix N, Lauridsen LLB, Olsen J, Parner ET, Liew Z, et al. Acetaminophen (paracetamol) exposure during pregnancy and pubertal development in boys and girls from a nationwide puberty cohort. *Am J Epidemiol.* 2019;188:34-46.
11. Panchaud A, Rousson V, Vial T, Bernard N, Baud D, Amar E, et al. Pregnancy outcomes in women on metformin for diabetes or other indications among those seeking teratology information services. *Br J Clin Pharmacol.* 2018;84:568-78.
12. Pasternak B, Svanstrom H, Hviid A. Ondansetron in pregnancy and risk of adverse fetal outcomes. *N Engl J Med.* 2013;368:814-23.
13. Norby U, Kallen K, Shemeikka T, Korkmaz S, Winbladh B. Pregnant women's view on the swedish internet resource drugs and birth defects intended for health care professionals. *Acta Obstet Gynecol Scand.* 2015;94:960-8.
14. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *Int J Methods Psychiatr Res.* 2011;20:40-9.
15. Rubin DB. *Multiple imputation for nonresponse in surveys*. Wiley, 1987.
16. Rubin DB. Inference and missing data. *Biometrika.* 1976;63:581-92.
17. Beau AB, Montastruc JL, Lacroix I, Montastruc F, Hurault-Delarue C, Damase-Michel C. Atropinic burden of drugs during pregnancy and psychological development of children: A cohort study in the efemeris database. *Br J Clin Pharmacol.* 2016;82:478-86.

18. Coomarasamy A, Devall AJ, Cheed V, Harb H, Middleton LJ, Gallos ID, et al. A randomized trial of progesterone in women with bleeding in early pregnancy. *N Engl J Med.* 2019;380:1815-24.
19. Magnus MC, Karlstad O, Haberg SE, Nafstad P, Davey Smith G, Nystad W. Prenatal and infant paracetamol exposure and development of asthma: The norwegian mother and child cohort study. *Int J Epidemiol.* 2016;45:512-22.
20. Radojcic MR, El Marroun H, Miljkovic B, Stricker BHC, Jaddoe VWV, Verhulst FC, et al. Prenatal exposure to anxiolytic and hypnotic medication in relation to behavioral problems in childhood: A population-based cohort study. *Neurotoxicol Teratol.* 2017;61:58-65.
21. Scherneck S, Schlinke N, Beck E, Grupe K, Weber-Schoendorfer C, Schaefer C. Pregnancy outcome after first-trimester exposure to metformin: A prospective cohort study. *Reprod Toxicol.* 2018;81:79-83.
22. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* 2019;110:63-73.
23. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf.* 2010;19:618-26.
24. Petersen I, Welch CA, Nazareth I, Walters K, Marston L, Morris RW, et al. Health indicator recording in uk primary care electronic health records: Key implications for handling missing data. *Clin Epidemiol.* 2019;11:157-67.
25. Lupattelli A, Wood M, Lapane K, Spigset O, Nordeng H. Risk of preeclampsia after gestational exposure to selective serotonin reuptake inhibitors and other antidepressants: A study from the norwegian mother and child cohort study. *Pharmacoepidemiol Drug Saf.* 2017;26:1266-76.
26. Strand BH, Dalgard OS, Tambs K, Rognerud M. Measuring the mental health status of the norwegian population: A comparison of the instruments scl-25, scl-10, scl-5 and mhi-5 (sf-36). *Nord J Psychiatry.* 2003;57:113-8.
27. Fink P, Orbol E, Hansen MS, Sondergaard L, De Jonge P. Detecting mental disorders in general hospitals by the scl-8 scale. *J Psychosom Res.* 2004;56:371-5.
28. Moodie EE, Delaney JA, Lefebvre G, Platt RW. Missing confounding data in marginal structural models: A comparison of inverse probability weighting and multiple imputation. *Int J Biostat.* 2008;4:Article 13.
29. Mojaverian N, Moodie EE, Bliu A, Klein MB. The impact of sparse follow-up on marginal structural models for time-to-event data. *Am J Epidemiol.* 2015;182:1047-55.

Acknowledgements

The authors acknowledge Prof Kate Lapane and Dr Shao-Hsien Liu for the fruitful discussions on missing data methods in observational studies.

Conflict of Interest statement

The authors declare no conflict of interest.

Introduction

Missing data is a global problem in human subjects research, and a serious threat to both validity and efficiency in effect estimation. Epidemiological studies on the safety of medication during pregnancy are all susceptible to missing data, i.e. data intended to be collected to answer a specific research question, but were not.[1] Missing data are ubiquitous and this is irrespective of the data source used, whether it is a healthcare database, a registry, a prospective birth cohort, or a clinical study.

The term "missing data" is frequently used to describe data that are explicitly missing, that is information that should have been recorded but for some reason was not: e.g., body weight at a pre-pregnancy primary care visit, or self-reported smoking status in birth registration. However, the term Missing data are ubiquitous and this is irrespective of the data source used, whether it is a healthcare database, a registry, a prospective birth cohort, or a clinical study.

Understanding threats to validity arising from missing data first requires an agreement on definitions. The term "missing data" has beenis also used to cover a wide array of possible scenarios where data are imperceptibly missing. For instance, c. Indeed, causal inference has been described as a missing data problem, in which only one of two potential outcomes has been observed.,.[2] Similarly, Somewhat less broadly, it is possible to describe classic epidemiologic biases, including information bias, confounding bias, and selection bias, can be framed as missing data problems since these are often described using potential outcomes.[3, 4] Although all these scenarios can plausibly describe missing data in studies of medication safety during pregnancy, Finally, a more narrow definition is frequently used to describe data that should have been recorded but for some reason was not: e.g., body weight at a pre-pregnancy primary care

~~visit, or self-reported smoking status in birth registration. All of these definitions can plausibly describe missing data in studies of medication safety during pregnancy, and all imply a set of analytic tools available to researchers. I~~in this paper~~, we will focus our attention on the latter definition~~explicit "missing data"~~, ~~in response to calls for increased use of methods for dealing with this kind of missing data problem.~~[5]

~~As advocated by the~~The CONSORT 2010 Statement[6] advocates ~~and thereby applicable to medication in pregnancy research, a~~ transparent reporting of the extent of missing data and how this issue was dealt with in the analysis~~, since that~~as this is crucial for readers to~~a~~ critically evaluate~~appraisal of~~ the study findings and ~~of~~ potential ~~risk of~~ biases. Recognition of the threat from these biases has resulted in calls for increased use of methods for dealing with missing data.[5] However, barriers exist that prevent applied pharmacoepidemiology researchers from assessing the potential gains to their own work, including understanding scenarios when simpler methods might be sufficient, or when complex approaches are needed. These barriers include a lack of resources that integrate missing data terminology and approaches with epidemiologic concepts, and a discussion of the strengths and weaknesses of the most common approaches.

We review the critical concepts for missing data problems, with the aim of integrating more traditional statistical language on missingness mechanisms with epidemiologic methods based on causal diagrams.[7] We have framed this commentary using examples from perinatal pharmacoepidemiology, including an applied ~~motivating~~ example from the Norwegian Mother and Child Birth Cohort (M~~O~~oBa): evaluating the effect of prenatal use of selective serotonin reuptake inhibitor~~s~~ (SSRI~~s~~) antidepressants on preeclampsia in the presence of missing data on relevant confounders such as smoking status in gestation.~~, when data on an important confounder- smoking- is missing.~~

Indeed, bias can arise when incomplete information on confounders, outcome measures, pregnancy duration, or even cohort selection criteria, are used to estimate prenatal exposure effects that would be obtained from the fully observed data, if these were available for each mother-child dyad.[3, 8-10]

Despite the advances in analyzing incomplete data via multiple imputation, and the well-established shortcomings of simpler approaches such as complete-case analysis or indicator variable for missingness,[10-12] missing data remain overlooked in a substantial number of epidemiological studies.[1, 13] Because causal diagrams have been proposed as a guiding framework to handle missing data,[7] it is of interest to explore their application in the context of medication in pregnancy research, and not least in evaluating selection bias from missing non-live births when birth cohorts are conditioned on live birth.

In this commentary we briefly describe general missing data mechanisms and methods, and illustrate missing data handling in recent medication in pregnancy research by data source utilized. We further present one motivating example on missing data analysis within the Norwegian Mother and Child Cohort Study (MoBa), a nationwide population-based cohort study where pregnant women completed three prenatal and multiple postnatal questionnaires.[14, 15] Our motivating MoBa example deals specifically with missing data at the variable level (missing information in one question or one item in a scale); although of importance, missing data on individual level (no information about the woman at all), or on an occasion level (one of three questionnaire in pregnancy non-completed) are beyond the scope of this commentary. Using the MoBa study example, we finally illustrate how the causal diagram framework can be of aid in

assessing risk of bias due to missing data in medication in pregnancy research.**Missing data methods and mechanisms**

Missing data are generally classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR)[5, 11, 16] as briefly described below.

Missing completely at random (MCAR)

Under this scenario, there are no systematic differences between the missing and the observed values.[5, 11] For example, if unexperienced health care personnel forget to ask about smoking during pregnancy, information about smoking will be missing at random in the pregnant woman's medical chart. The same occurs when study participants randomly forget to fill in~~n~~ or skip responses. There is no risk of bias with MCAR data, but there will be loss of precision.

Missing at random (MAR)

Missing at random is classified as any systematic difference between the missing values and the observed values, which can be explained by the observed data.[5, 11] For instance, depressed pregnant women may be less likely to report smoking than non-depressed. ~~Yet, this pattern of missingness may be explained by observed data on socioeconomic status, age, or other factors. MAR occurs often in epidemiological studies, and is often also called informative missingness. Whenever data are assumed to be MAR, bias can arise, and this can be reduced by employing statistical methods to handle missing data.~~

Missing not at random (MNAR)

Missing not at random occurs in situations when systematic differences remain between the missing values and the observed values, even after the observed data are taken into account;

missingness is thus related to unmeasured variables. For example, women who smoke during pregnancy may less likely report their smoking status~~.~~. When missingness in a variable depends on the missing value itself, the unbiased estimate is not recoverable in observed data~~and likewise depressed women are less likely to show up at the study visit, and in both instances missingness unrelated to observed data. Whenever data are assumed to be MNAR, bias will arise, and unfortunately there is no adequate method to handle this problem.~~

Exploring extent and patterns of missingness

Although Little's test may help researchers to identify missingness that is MCAR vs. MAR, this test is not conclusive. In addition, n~~N~~o numerical diagnostics can differentiate MAR from MNAR. This means ~~,~~ we are left with logical reasoning to inform us on the mechanism behind data missingness ~~(See also DAG framework with missing data)~~. Exploring the extent and pattern of missing data in one's own data sample (for example by cross-tabulating variables with missing data against exposure and outcome), as well as using findings from previous studies and normative data (e.g. score distribution in a reference population) can give a hint of the underlying mechanism of missingness. This is important to appraise as it will guide decision making of missing data handling: the various approaches to missing data analysis require different assumptions about the underlying mechanisms.~~. Recent research has also shown that the proportion of missing data should not be the major driver for the decision on how to handle missing data,[17] In fact, even when the extent of missing data is large, results can still be unbiased provided that the MAR assumption is met and methods to handle missing data have been adequately applied.~~

Field Code Changed

Methods to handle missing data

Multiple methods for handling missing data are used in ~~medication in pregnancy~~perinatal pharmacoepidemiology research. These methods fall into two broad categories: analyze the observed data (complete case analysis), or use some principled method for filling in the missing data (imputation). ~~.~~In ~~the~~ complete case analysis (CCA), observations with missing data on ~~one or multiple~~relevant variables are dropped from the analysis. This approach will always produce unbiased results under the MCAR assumption, and may produce unbiased results under MAR or MNAR. ~~albeit it will lead to loss of precision due to reduced sample size. The complete case analysis requires enough complete cases to estimate the model and the assumption of MCAR data. Despite this strong assumption which is rarely the case in observational studies, this method~~CCA is commonly used in ~~medication in pregnancy research~~perinatal pharmacoepidemiology due to its simplicity (Table 1). In database linkage studies where study size is large, the loss of data has less impact on precision than in smaller size or different design studies.[18-22]

Single imputation comprises a set of techniques where missing value are replaced by a value from the observed data, for instance the mean or mode. The imputed values are assumed to be equal to the values that would have been observed if data had been complete. This method, however, underestimates uncertainty about the missing values and will therefore result in ~~too small~~ standard errors that are too small.[5, 11] In the study by Panchaud et al,[23] gestational age was conditionally imputed for 6% of the pregnancies based on the sample mean. In the study by Pasternak et al,[24] missing information on several baseline maternal characteristics was replace using the mode. In longitudinal studies with repeated variable measurement, for example using questionnaires at several time points in pregnancy, the "last value carried forward" technique can be used to replace missing values with the last measured value of the individual, as done by

Norby et al.[25] This method ~~makes the assumption~~assumes that the observation of the individual remains the same since the last measured observation. Due to ~~well established~~well-established shortcoming~~s~~,[5, 11] ~~—~~single imputation ~~s~~ techniques are less used in ~~medication in pregnancy research~~perinatal pharmacoepidemiology.

More advanced model-based methods for handling missing data have become more accessible to researchers in recent years through packages in standard statistical software. The two most common model-based methods are maximum likelihood using the ~~estimation~~ expectation maximization (EM) algorithm and multiple imputation.[12, 16, 26] These are considered model-based methods since the researcher must make assumptions about the joint distribution of all variables in the model (including both outcomes and predictors).

M~~The m~~aximum likelihood methods using the EM algorithm ~~does not fill in the missing values, but rather~~ uses each ~~case's~~ observation's available data to compute maximum likelihood estimates, rather than filling in the missing values.~~.~~ It runs until the algorithm converges to the "best fit" model for a set of data.

The multiple imputation (MI) method fills in missing values by averaging from the distribution of the missing data given the observed data in a way that accounts for the uncertainty associated with the missing values. In MI by chained equations (MICE) a series of regression models are run whereby each variable with missing data is modeled conditional upon the other variables in the data.[26] At the end of one cycle, all missing values have been replaced with predicted values (imputations). The process is repeated for a number of cycles, with the imputations being updated at each cycle, finally resulting in one imputed dataset. The number of imputed datasets is generally between 5 and 20. Standard errors are calculated using Rubin's rules.[10, 12] The MI

approach produces valid estimates under the MAR assumption. This is a weaker assumption than MCAR and more likely to hold in observational studies. MI is a computationally intensive method which is increasingly used in perinatal pharmacoepidemiology ~~medication in pregnancy~~ research (Table 1).[19, 27-31] Yet, this needs to be applied after careful reflection about the missing data to avoid misleading conclusions. For a comprehensive review of multiple imputation for missing data in epidemiological studies we recommend the papers by Sterne et al,[11] and Perkins et al.[5] ~~–~~Recent research has also shown that the proportion of missing data should not be the major driver for the decision on how to handle missing data.[17] In fact, even when the extent of missing data is large, results can still be unbiased provided that the MAR assumption is met and methods to handle missing data have been adequately applied.

Field Code Changed

Missing data approaches in recent medication in pregnancy literature

Table 1 summarizes the reporting and handling of missing data in recent perinatal pharmacoepidemiology~~medication in pregnancy~~ studies, by type of data source utilized. Of note, this study overview serves as common ground for appraising current methodological gaps, and it is not a comprehensive, systematic extract of the literature. Transparent reporting of the extent and handling of missing data, and the uptake of multiple imputation methods, ~~remain to date~~remains limited. For instance, in multiple cases we computed the extent of missing data in a study using baseline characteristic data of the study sample based on numbers reported in each manuscript; in some studies, it was unclear what missing data approach was used. The majority of studies reported missing data on confounding variables, in different extent (from <1% to 65%) depending on the data source utilized.

On the basis of the ~~adopted~~ missing data definition used by study authors, and the information reported, missing data do not seem to be a major problem in health registry, administrative claims, or pregnancy registry. This contrasts with~~, as opposed to~~ studies set in birth cohorts, teratology information services, or general practice databases, which often have to contend with much higher levels of missingness, and with patterns that are likely to be informative. The substantial problem of missing data in these ~~latter~~ study types has ~~however~~ promoted important methodological research on the topic,[32, 33] as well as a greater uptake of multiple imputation methods by researchers using this type of data (Table 1). Simpler approaches to handle missing data such as indicator variable, were not often reported in the papers we evaluated; this is encouraging given the well-established shortcomings of the method. Study authors rarely stated any ~~Although untestable, the~~ assumptions they made about the underlying mechanism of missingness ~~given the observed data was rarely available~~ in the ~~examined pregnancy~~ literature we reviewed.

DAG framework with missing data

Directed acyclic graphs (DAGs) can provide helpful insights into potential biases from assuming various missingness mechanisms. Figure 1 introduces a simplified causal model for the effect of prenatal SSRI exposure on preeclampsia. In this model, we assume a causal effect of depression severity on SSRI use and on smoking, and that smoking has an effect on preeclampsia risk. If these assumptions hold, we could estimate the effect of SSRI use on preeclampsia by conditioning on smoking and depression severity. If some fraction of the study sample lacks data on smoking, assumptions about the mechanism that explains the missingness will point to different strategies for analyzing our data. In Figure 1A, smoking is missing completely at random (MCAR), and we can fit a model for the effect of SSRI use on preeclampsia risk,

adjusting for depression severity and smoking, in the complete case sample only, without risk of bias. Figures 1B shows that if missingness in smoking status is explained by depression severity, we can also estimate unbiased effects in the complete case sample, as the covariates required for confounding control also block bias paths from missingness to the outcome. For missing data mechanisms where the missingness is predicted by the missing values, as in Figure 1C, or when the probability of being a complete case depends on the outcome, as in Figures 1D and 1E, complete case analysis will result in a biased estimate. Finally, the presence of an auxiliary variable (that is, a variable that predicts missingness but is unrelated to the causal mechanism being considered) allows for unbiased and efficient effect estimation via multiple imputation.

~~Motivating~~ Applied example: prenatal antidepressant use and risk of preeclampsia

~~We present as~~As a motivating~~n~~ example, we present recent work on the association between use of selective serotonin re-uptake inhibitor (SSRI) antidepressants during gestation and risk of late-onset preeclampsia, using data from ~~the Norwegian Mother and Child (~~the MoBa~~)~~ cohort study.[34] MoBa is a nation-wide, population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health, with recruitment occurring between ~~in~~ 1999-2008.[23] Pregnant women were recruited from all over Norway at the time of their routine ultrasound at 17-18 weeks of gestation. Data were gathered prospectively by self-administered questionnaires. The cohort now includes 114500 children, 95200 mothers and 77300 fathers, all of whom are followed as long as they continue to participate in the study~~.~~[23] MoBa has a license from the Norwegian Data Inspectorate and approval from The Regional Committee for Medical Research Ethics. All participants gave their written informed consent prior to participation.

In our study, we first explored patterns of missing data on important confounders by exposure and outcome strata. Missing values on these confounders ranged from 1-3% for maternal smoking and body mass index, to 7-8% for education and weight gain. Missing information on maternal depressive and anxiety symptom severity in pregnancy, measured via the Hopkins Symptom Checklist-25 (SCL-25) at gestational week 17 (5 items, SCL-5) and 30 (8 items, SCL-8),[35, 36] was as follows: 5% and 10% on at least one of the SCL-5 or SCL-8 items, respectively; 15% total missing information simultaneously on either scale. However, only few women (< 3%) completed none or less than a half of the items composing the individual SCL scales. The missing data mechanism in our study seemed to be linked to maternal age and to the extent of completion of the SCL items, but importantly, it did not seem to be associated with the outcome, late-onset preeclampsia. ~~.~~ Based on this and under the MAR assumption,[34] we conducted three sets of analyses: i) complete case analysis; ii) multiple imputation of missing data on the two SCL scales only (approach I); and iii) multiple imputation of missing data on the two SCL scales and on other maternal confounders (approach II). As shown in Table 2, the adjusted and weighted association measures were higher and less precise in the complete case analysis than in the other two sets. However, the results of the complete case analysis expanded to pregnancies with only SCL imputed values (approach I) were similar to those obtained in the fully imputed models (approach II). Increasing sample size and higher statistical power following multiple imputation can indeed explain these discrepancies. The extent of missing data on confounders other than the SCL between the complete-case and approach I (31.9% vs 24.1%) analysis was however not ~~so~~ substantial. Hence, because in this example missing data seemed to relate to the extent of completion of the SCL items, we could not exclude the possibility that a complete case

analysis approach would yield biased estimates.[11, 37, 38] In the context of this applied example, results from approach II were thereby considered as those least biased. [11, 37, 38]

~~**DAG framework with missing data**~~

Implications for applied researchers

Formatted: Left

Methods for identifying, analyzing, and mitigating bias from missing data have advanced significantly in recent years, and are seeing greater uptake in ~~the~~ applied ~~research literature on medication use in pregnancy~~perinatal pharmacoepidemiology research. Based on our survey of the literature, we have several recommendations for applied researchers who need to analyze data with missing values. These recommendations are made bearing in mind that there is no missing data handling solution that fits all research contexts.

First: where possible, limit missingness during collection of data. Recognize that no statistical method can make up for careful study design and data curation. Sometimes the assumptions a specific case of missing data require are simply so unrealistic that the effect estimate is unlikely to be informative. *Second*: carefully diagnose missingness, and use subject-area knowledge as well as exploratory and descriptive data analysis to understand plausible mechanisms of missingness. We suggest that a minimum standard for missing data analysis should be a complete reporting of missingness within strata of exposure and outcome. Researchers should consider the use of causal graphs to make their assumptions about missingness mechanisms explicit. ~~–~~ *Third*: Be aware that the proportion of missing data should not be the major driver for the decision on how to handle missing data, but rather the assumed mechanism as to why data

are missing. ~~*Third*~~*Fourth*: include a statistical analyst with expertise in missing data methods. Inappropriate analyses using these complex methods can result in seriously biased results. *Finally*: apply strategies for missing data mitigation under different assumptions, and include evaluations of robustness results under these assumptions. For example, including both the complete case analysis ~~(under an MCAR assumption)~~ and the multiply imputed results ~~(under an MAR assumption)~~ can allow readers to decide which estimate they prefer, depending on assumptions about the missingness mechanism.

C~~Greater and c~~areful attention to missing data, and to the assumptions required for analysis of missing data, is necessary in ~~medication in pregnancy safety research~~all areas of research, including perinatal pharmacoepidemiology. With transparent reporting of the extent and assumed mechanisms of missing data, and by applying strategies for missing data mitigation under different assumptions, future research can avoid the problems that result from failure to consider this important source of bias.

Figure Legends

Figure 1. ~~Four directed acyclic graph-based examples of missingness.~~Causal diagrams showing relationships between prenatal SSRI use, maternal depression severity, preeclampsia, and smoking, as well as a binary indicator, $Miss_{SMK}$, denoting missing information in the smoking variable.

References

1. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials.* 2004;1:368-76.
2. Hernan MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health.* 2004;58:265-71.
3. Howe CJ, Cain LE, Hogan JW. Are all biases missing data problems? *Curr Epidemiol Rep.* 2015;2:162-71.
4. Edwards JK, Cole SR, Westreich D. All your data are always missing: Incorporating bias due to measurement error into the potential outcomes framework. *Int J Epidemiol.* 2015;44:1452-9.
5. Perkins NJ, Cole SR, Harel O, Tchetgen Tchetgen EJ, Sun B, Mitchell EM, et al. Principled approaches to missing data in epidemiologic studies. *Am J Epidemiol.* 2018;187:568-75.
6. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340.
7. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. *Stat Methods Med Res.* 2012;21:243-56.
8. Eberg M, Platt RW, Filion KB. The estimation of gestational age at birth in database studies. *Epidemiology.* 2017;28:854-62.
9. Bengtson AM, Westreich D, Musonda P, Pettifor A, Chibwesha C, Chi BH, et al. Multiple overimputation to address missing data and measurement error: Application to hiv treatment during pregnancy and pregnancy outcomes. *Epidemiology.* 2016;27:642-50.
10. Rubin DB. Inference and missing data. *Biometrika.* 1976;63:581-92.
11. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ.* 2009;338:b2393.
12. Rubin DB. *Multiple imputation for nonresponse in surveys*. Wiley, 1987.
13. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol.* 2008;168:355-7.
14. Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C, et al. Cohort profile: The norwegian mother and child cohort study (moba). *Int J Epidemiol.* 2006;35:1146-50.
15. Magnus P, Birke C, Vejrup K, Haugan A, Alsaker E, Daltveit AK, et al. Cohort profile update: The norwegian mother and child cohort study (moba). *Int J Epidemiol.* 2016;45:382-8.
16. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *J R Stat Soc Series B Stat Methodol.* 1977;39:1-38.
17. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* 2019;110:63-73.
18. Bateman BT, Heide-Jorgensen U, Einarsdottir K, Engeland A, Furu K, Gissler M, et al. Beta-blocker use in pregnancy and the risk for congenital malformations: An international cohort study. *Ann Intern Med.* 2018;169:665-73.
19. Caniglia EC, Patel K, Huo Y, Williams PL, Kapetanovic S, Rich KC, et al. Atazanavir exposure in utero and neurodevelopment in infants: A comparative safety study. *AIDS.* 2016;30:1267-78.
20. Dhalwani NN, Szatkowski L, Coleman T, Fiaschi L, Tata LJ. Stillbirth among women prescribed nicotine replacement therapy in pregnancy: Analysis of a large uk pregnancy cohort. *Nicotine Tob Res.* 2019;21:409-15.

Field Code Changed

Formatted: Norwegian (Bokmål)

Formatted: Norwegian (Bokmål)

Formatted: Norwegian (Bokmål)

21. Elkjaer LS, Bech BH, Sun Y, Laursen TM, Christensen J. Association between prenatal valproate exposure and performance on standardized language and mathematics tests in school-aged children. *JAMA Neurol.* 2018;75:663-71.
22. Ernst A, Brix N, Lauridsen LLB, Olsen J, Parner ET, Liew Z, et al. Acetaminophen (paracetamol) exposure during pregnancy and pubertal development in boys and girls from a nationwide puberty cohort. *Am J Epidemiol.* 2019;188:34-46.
23. Panchaud A, Rousson V, Vial T, Bernard N, Baud D, Amar E, et al. Pregnancy outcomes in women on metformin for diabetes or other indications among those seeking teratology information services. *Br J Clin Pharmacol.* 2018;84:568-78.
24. Pasternak B, Svanstrom H, Hviid A. Ondansetron in pregnancy and risk of adverse fetal outcomes. *N Engl J Med.* 2013;368:814-23.
25. Norby U, Kallen K, Shemeikka T, Korkmaz S, Winbladh B. Pregnant women's view on the swedish internet resource drugs and birth defects intended for health care professionals. *Acta Obstet Gynecol Scand.* 2015;94:960-8.
26. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *Int J Methods Psychiatr Res.* 2011;20:40-9.
27. Beau AB, Montastruc JL, Lacroix I, Montastruc F, Hurault-Delarue C, Damase-Michel C. Atropinic burden of drugs during pregnancy and psychological development of children: A cohort study in the efemeris database. *Br J Clin Pharmacol.* 2016;82:478-86.
28. Coomarasamy A, Devall AJ, Cheed V, Harb H, Middleton LJ, Gallos ID, et al. A randomized trial of progesterone in women with bleeding in early pregnancy. *N Engl J Med.* 2019;380:1815-24.
29. Magnus MC, Karlstad O, Haberg SE, Nafstad P, Davey Smith G, Nystad W. Prenatal and infant paracetamol exposure and development of asthma: The norwegian mother and child cohort study. *Int J Epidemiol.* 2016;45:512-22.
30. Radojcic MR, El Marroun H, Miljkovic B, Stricker BHC, Jaddoe VWV, Verhulst FC, et al. Prenatal exposure to anxiolytic and hypnotic medication in relation to behavioral problems in childhood: A population-based cohort study. *Neurotoxicol Teratol.* 2017;61:58-65.
31. Scherneck S, Schlinke N, Beck E, Grupe K, Weber-Schoendorfer C, Schaefer C. Pregnancy outcome after first-trimester exposure to metformin: A prospective cohort study. *Reprod Toxicol.* 2018;81:79-83.
32. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf.* 2010;19:618-26.
33. Petersen I, Welch CA, Nazareth I, Walters K, Marston L, Morris RW, et al. Health indicator recording in uk primary care electronic health records: Key implications for handling missing data. *Clin Epidemiol.* 2019;11:157-67.
34. Lupattelli A, Wood M, Lapane K, Spigset O, Nordeng H. Risk of preeclampsia after gestational exposure to selective serotonin reuptake inhibitors and other antidepressants: A study from the norwegian mother and child cohort study. *Pharmacoepidemiol Drug Saf.* 2017;26:1266-76.
35. Strand BH, Dalgard OS, Tambs K, Rognerud M. Measuring the mental health status of the norwegian population: A comparison of the instruments scl-25, scl-10, scl-5 and mhi-5 (sf-36). *Nord J Psychiatry.* 2003;57:113-8.
36. Fink P, Orbol E, Hansen MS, Sondergaard L, De Jonge P. Detecting mental disorders in general hospitals by the scl-8 scale. *J Psychosom Res.* 2004;56:371-5.
37. Moodie EE, Delaney JA, Lefebvre G, Platt RW. Missing confounding data in marginal structural models: A comparison of inverse probability weighting and multiple imputation. *Int J Biostat.* 2008;4:Article 13.

Formatted: Norwegian (Bokmål)

38. Mojaverian N, Moodie EE, Bliu A, Klein MB. The impact of sparse follow-up on marginal structural models for time-to-event data. *Am J Epidemiol.* 2015;182:1047-55.

Formatted: Font: (Default) Times New Roman, 12 pt

Acknowledgements

The authors acknowledge Prof Kate Lapane and Dr Shao-Hsien Liu for the fruitful discussions on missing data methods in observational studies.

Conflict of Interest statement

The authors declare no conflict of interest.

Table 1: Examples of functional approaches to missing data in recent pregnancy medication safety research, by type of data source

Study	Medication, outcome	Diagnosing missing data		Handling and analyzing missing data[a]				
		Mechanism assumption	Amount, variable type	Complete case	Single imputation	Multiple imputation	Indicator variable	Others
Birth cohorts								
Magnus et al.[27]	Paracetamol, child asthma	Unspecified	10-15% multiple confounders	SA		MA		
Radojcic et al.[28]	Anxiolytics/ hypnotics, child development	Unspecified	<1-18% multiple confounders			MA		
Ernst et al.[20]	Paracetamol, child puberty	Unspecified	<5% multiple confounders	MA[b]				
Caniglia et al.[17]	Atazanavir, child development	Unspecified	40% outcome 1-40% multiple confounders	MA		SA		
Brandlistuen et al.[37]	Paracetamol, child development	Unspecified	<1-4% multiple outcomes					MA[c]
			<1-18% multiple confounders				MA	
Case-control studies (birth defects surveillance)								
Tinker et al.[38]	Anxiolytics, congenital anomaly	Unspecified	<10% multiple confounders[d]	MA[b]				
Health registries and administrative claims								
Pasternak et al.[22]	Ondansetron, adverse fetal outcomes	Unspecified	<1-7% multiple confounders		MA			
Beau et al.[25]	Atropinic drugs, child development	MAR	5-19% multiple confounders			MA		
Bateman et al.[16]	β Blocker, congenital anomaly	Unspecified	1% parity	MA				
Elkjaer et al.[19]	Valproate, child cognition	Unspecified	<1% multiple confounders	MA				

Study	Medication, outcome	Diagnosing missing data		Handling and analyzing missing data[a]				
		Mechanism assumption	Amount, variable type	Complete case	Single imputation	Multiple imputation	Indicator variable	Others
General practice research databases								
McGrogan et al.[39]	Statins, pregnancy loss	Unspecified	5-16% multiple confounders				MA	
Dhalwani et al.[18]	Nicotine replacement, stillbirth	Unspecified	8-30%	MA[b]			MA[b]	
Teratology Information Service based studies								
Panchaud et al.[21]	Metformin, adverse pregnancy outcomes	Unspecified	6% gestational age		MA			
			57% body mass index				MA	
Scherneck et al.[29]	Metformin, congenital anomaly and spontaneous abortion	MAR	<1-43% multiple confounders[d]			MA		
Pregnancy registries								
Cohen et al.[40]	Quetiapine, congenital anomaly	Unspecified	Unspecified	MA				
Randomized clinical trials								
Coomarasamy et al.[26]	Progesterone, live-birth and other neonatal outcomes	Unspecified	3% outcome	MA		SA		

Abbreviations: MA=Main analysis; MAR=Missing at Random; SA=Sensitivity analysis.

[a]Inverse Probability Weighting (IPW) method not presented since we found no study using it.

[b]Unspecified in the article how missing data were handled; we suppose that to be a complete case approach or an indicator variable use.

[c]Other method used: Expectation maximization.

[d]The extent of missing data was unspecified in the original study; we computed that by using data reported in the baseline factor table or in the text.

Table 2: Association of prenatal SSRI exposure with maternal risk of late-preeclampsia under three missing data handling scenarios

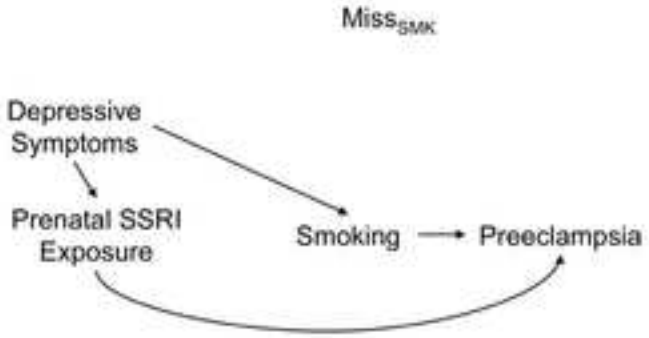
	Complete-case[a] (n=3913)	Multiple Imputed, I[b] (n=4361)	Multiple Imputed, II[c] (n=5745)	Complete-case[a] (n=3913)	Multiple Imputed, I[b] (n=4361)	Multiple Imputed, II[c] (n=5745)
	Adjusted analysis (unweighted)			Weighted analysis		
	Adjusted RR (95% CI)	Adjusted RR (95% CI)	Adjusted RR (95% CI)	Weighted RR (95% CI)	Weighted RR (95% CI)	Weighted RR (95% CI)
SSRI, early pregnancy	1.22 (0.76-1.94)	1.07 (0.67-1.69)	0.96 (0.63-1.46)			
SSRI, midpregnancy	1.04 (0.49-2.21)	0.92 (0.43-1.95)	0.92 (0.48-1.79)	0.48 (0.21-1.11)	0.63 (0.30-1.32)	0.66 (0.33-1.28)
SSRI, late pregnancy	1.16 (0.55-2.46)	1.03 (0.49-2.17)	1.08 (0.57-2.07)	2.28 (0.88-5.87)	1.52 (0.65-3.56)	1.34 (0.61-2.93)
SSRI, any time	1.26 (0.80-1.98)	1.09 (0.70-1.71)	0.96 (0.64-1.45)			

Reference: unexposed pregnancies in the corresponding time window. The weighted analyses correspond to Marginal Structural Model with inverse probability of treatment weight. The adjusted analyses correspond to multivariate modified-Poisson regression models.

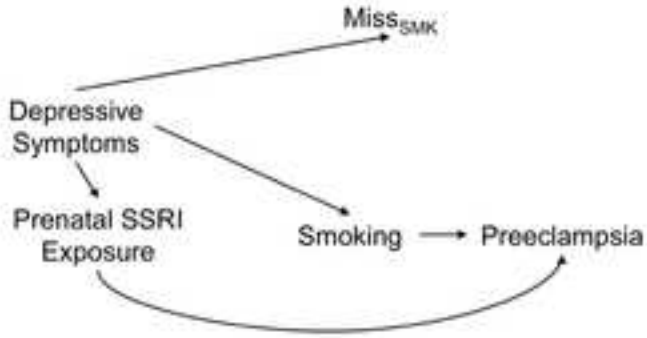
[a]Including only observations with complete data on all confounders; [b]including observations where multiple imputation was done for missing data on the two SCL scales only; [c]including observations where multiple imputation was done for missing data on the two SCL scales as well as on other maternal confounders.

Figure

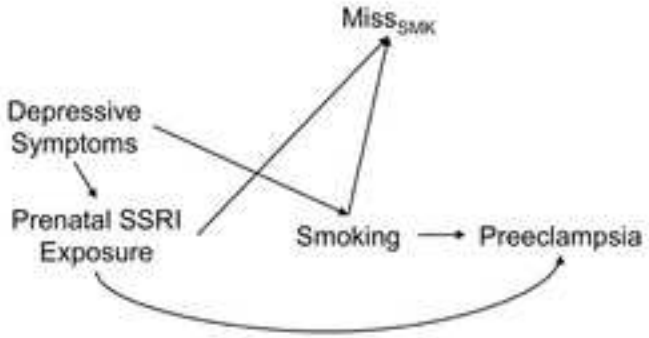
A. Missing Completely At Random (MCAR)



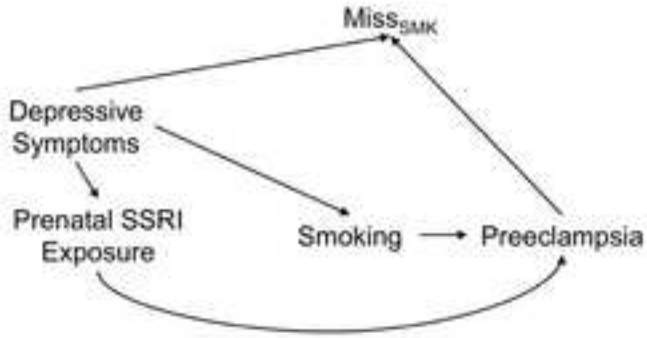
B. Missing At Random (MAR) depending on a confounder



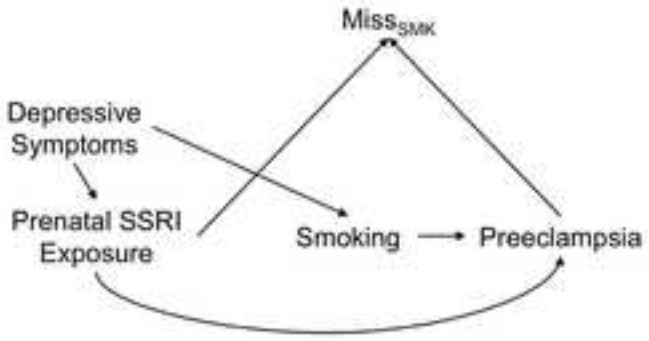
C. Missing Not At Random (MNAR) depending on exposure and confounder



D. MAR depending on outcome and confounder



E. MNAR depending on outcome and exposure



F. MAR depending on a confounder; auxiliary variable that explains missingness