
Peer reviewed version

Link to published version (if available): 10.1002/pds.4336

Link to publication record in Explore Bristol Research

PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at https://onlinelibrary.wiley.com/doi/abs/10.1002/pds.4336. Please refer to any applicable terms of use of the publisher.

**University of Bristol - Explore Bristol Research**

**General rights**

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/
# Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control

**Journal:** *Pharmacoepidemiology and Drug Safety*

**Manuscript ID:** PDS-17-0079.R1

**Wiley - Manuscript type:** Review

**Date Submitted by the Author:** n/a

**Complete List of Authors:**
- Wood, Mollie; University of Oslo, Department of Pharmacy
- Lapane, Kate; University of Massachusetts Medical School, Quantitative Health Sciences
- van Gelder, Marleen; Radboud Universiteit, Radboud REshape Innovation Center
- Rai, Dheeraj; University of Bristol, Psychiatry
- Nordeng, Hedvig; University of Oslo, Department of Pharmacy

**Keywords:** methods review, confounding bias, pregnancy safety

**Abstract:**
Understanding the safety of medication use during pregnancy relies on observational studies: however, confounding in observational studies poses a threat to the validity of estimates obtained from observational data. Newer methods, such as marginal structural models and propensity calibration, have emerged to deal with complex confounding problems, but these methods have seen limited uptake in the pregnancy medication literature. In this article, we provide an overview of newer advanced methods for confounding control and show how these methods are relevant for pregnancy medication safety studies.
Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control.

Mollie E. Wood¹, Kate L. Lapane², Marleen M.H.J. van Gelder³,⁴, Dheeraj Rai⁵, Hedvig Nordeng¹,⁶

Affiliations:
1. PharmacoEpidemiology and Drug Safety Research Group, School of Pharmacy, University of Oslo, Norway
2. Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States
3. Department for Health Evidence, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, The Netherlands
4. Radboud REshape Innovation Center, Radboud university medical center, Nijmegen, The Netherlands
5. School of Social and Community Medicine, University of Bristol, United Kingdom
6. Department of Child Mental and Physical Health, Norwegian Institute of Public Health, Oslo, Norway

Current word count (everything, incl. references): 5974
References: 54
Figures: 2
Tables: 1
Supplemental Material: Box 1, Figure S1
Abstract.
Understanding the safety of medication use during pregnancy relies on observational studies; however, confounding in observational studies poses a threat to the validity of estimates obtained from observational data. Newer methods, such as marginal structural models and propensity calibration, have emerged to deal with complex confounding problems, but these methods have seen limited uptake in the pregnancy medication literature. In this article, we provide an overview of newer advanced methods for confounding control and show how these methods are relevant for pregnancy medication safety studies.

Key Points.
- Studies of the safety of medication use during pregnancy depend mainly on observational studies, which are subject to confounding bias.
- Novel methods for confounding control have seen limited uptake in the pregnancy medication safety literature.
- Application of novel methods is necessary to appropriately address the complex confounding scenarios found in pregnancy studies.
More than half of all pregnant women in the US and other western countries take medication during pregnancy, making studies of the safety and efficacy of medications a pressing public health concern. Studying medication safety in pregnancy presents particular challenges: effects of medication exposure on fetal development can be unpredictable, vulnerability to exposure changes during pregnancy, and outcomes of interest may occur early in fetal development but be detected later, as is the case for birth defects. In the general population, knowledge of medication efficacy and safety is primarily based on randomized controlled trials. However, randomized trials routinely exclude pregnant women due to uncertainties about the effects of medications on fetal development, meaning that studies of medication safety in pregnancy must rely on reproductive toxicity studies in animals and on observational data in humans. Several landmark cases have taught us that animal models for teratogenicity do not necessarily translate to humans, as was the case in the thalidomide disaster. Observational studies, which can include data sources such as pregnancy and birth cohorts, administrative and registry data, and pregnancy and birth registries, represent important opportunities for understanding the risks and benefits for medication use in pregnancy, and in 2005, the Food and Drug Administration (FDA) acknowledged that observational studies are the best method for assessing the maternal and fetal safety of using medication during pregnancy. However, confounding is a major source of bias in observational studies, and as such, it is the focus of this paper. Recent years have seen the rapid development of advanced methods for dealing with confounding; yet, uptake of these methods has been slow in the pregnancy medication literature. This is unfortunate, because in the pregnancy medication research field, it is arguably especially important that researchers make use of the best methods for confounding control and causal inference, because data from randomized controlled trials are unlikely to be available, and the consequences for getting the wrong answer are so profound: failing to detect true effects of medication exposure can have enormous effects in the population, and falsely raising the alarm for a safe drug can result in women forgoing needed therapies and in some cases, terminating wanted pregnancies.
In this paper, we advocate for a greater use of advanced methods for confounding control in the pregnancy medication safety research field, and provide an overview of these methods under the following framework:

(1) How does this method help us to make fair comparisons between the exposed and unexposed groups?
(2) How has this method been applied in the pregnancy medication literature?
(3) How is the method used in practice?
(4) What are the important assumptions for this method?
(5) What are the major strengths and limitations of the method?

Table 1 provides an outline of pregnancy medication studies using advanced methods to deal with confounding. This paper gives a useful reference for both students and experienced researchers who wish to gain new skills in advanced methods for confounding control.

Confounders and Causal Inference

First, a definition of confounding: a confounder is a factor in a study that (a) is associated with (and precedes) the medication exposure, (b) is a risk factor for the outcome, and (c) does not lie on the causal pathway between medication use and outcome. In randomized studies, we can be reasonably confident that confounders (both measured and unmeasured) are balanced between exposed and unexposed groups, whereas in observational studies, this is almost certainly not the case. If the distribution of confounders is different for the exposed and unexposed groups, estimates of effect may be biased. Taking confounding into account is essential in observational studies. Further, it is important for researchers to carefully consider the timing of the confounder relative to the exposure and outcome.⁴

Second, we focus on the definition of causal effects arising from the counterfactual: that is, we would like to understand what the experience of women who took a medication during pregnancy would have been, had she not taken the medication. In this definition, confounding exists when women who did not take the medication of interest are not adequate stand-ins for the women who did. The counterfactual approach to confounding clarifies a concept that researchers understand intuitively: that we must
strive to make fair comparisons between exposed and unexposed groups. For example, for researchers studying the effects of antidepressant exposure on congenital malformations, comparing a group of women using antidepressants who also have severe depression to a group of women with no antidepressant use and no history of depression would be an unfair comparison: the unexposed group is not an adequate stand-in for the experience of women using antidepressants.

**Confounding in pregnancy medication studies**

Application of advanced confounding control methods begins with a thorough review of the literature and consultation with subject-area experts. Directed acyclic graphs (DAGs) provide a graphical means to represent the causal structure the investigator believes is present, and guide study design, data collection, and analysis. Figure 1 is an example DAG showing one possible causal model for prenatal antidepressant exposure and childhood neurodevelopment. This DAG shows potential biasing paths, including confounders (other psychiatric illness, other psychiatric medication use, depression severity, and genetics) which should be controlled as far as possible, as well as a mediator (gestational age), and a collider (live birth). Several non-biasing paths, including a risk factor for the outcome that is unrelated to the exposure (child gender) and a predictor of exposure that is unrelated to the outcome (pre-pregnancy antidepressant use) are also shown. Obtaining unbiased effect estimates requires investigators to identify and control confounding, insofar as it is possible to do so, while avoiding bias from inappropriate control for colliders and mediators and loss of precision or confusing interpretation of estimates arising from control for factors only related to the exposure or outcome but not both.

Specific types of confounders arise often enough that they are grouped together. *Confounding by indication* challenges the researcher to distinguish whether the outcome of interest is caused by the drug under study or the disorder being treated. Including a disease comparison group (women with the same disease and ideally the same disease severity, but not treated with the drug), or comparing drug use across different indications, offers advantages over studies comparing exposed cases to healthy controls only. *Confounding by concomitant medication use* occurs when women use multiple
medications, and it is the concomitant medication, not the index drug, that is responsible for the observed effects. Other potential confounders include maternal factors such as lifestyle, paternal characteristics, genetic risk, and complex familial factors like home environment. These confounders may be measured with varying degrees of completeness or quality, depending on the data source, study design, and difficulty of measurement. We have grouped the available advanced methods for confounding control into categories based on qualities of confounder data, including whether and when it was measured.

I. Methods for measured confounders

In Box 1 (supplemental material), we include a simplified illustration of confounding by measured factors and the methods to address confounding.

Confounder summary scores and marginal structural models are useful for measured confounders. These methods reduce a large amount of information about an individual into a single summary score: in short, two individuals can have the same summary score but different individual confounder values (e.g., a woman with a predicted probability of taking an antidepressant during pregnancy of 0.5 might be an unemployed smoker with anxiety, or a non-smoking lawyer with depression), but because their distribution of confounders is equivalent, any differences in outcome will be attributable only to exposure to the drug of interest. Fair comparisons between exposure groups can then be made because within each stratum of exposure, the distribution of common causes of exposure and outcome are the same.

I.A. Propensity Scores (and other confounder summary scores)

Studies of medication safety in pregnancy often require consideration of, and adjustment for, a large number of confounders. The propensity score, which is the probability of exposure given observed confounders, reduces the set of confounders to a single confounder summary score. Propensity scores are more commonly used in the medical literature; however, other confounder summary score methods, including disease risk scores (preferred in the case of rare exposures) and polygenic risk scores (useful for cases when genetic confounding is of high concern) are available.
Propensity scores are typically constructed using multivariable logistic regression, where exposure is the dependent variable and confounders are the independent variables. The propensity score model should include variables that are confounders or predictors of the outcome; inclusion of factors that are only predictors of exposure will increase variance without decreasing bias. High dimensional propensity scores, which include thousands of variables identified through computational algorithms, may also be useful for adjusting for unmeasured confounders, if the measured variables are partial proxies for the unmeasured confounders. The propensity score can be used to match, stratify, adjust, or weight the outcome model. Propensity scores, including high dimensional propensity scores, have seen increased uptake in the pregnancy literature, i.e. safety studies on ondansetron, lithium, antidepressants, and statins in pregnancy, but their use is still minimal compared to multivariable regression (Table 1). Box 1, in the supplemental material, gives a simplified explanation of propensity score matching and weighting.

Assumptions: Use of propensity scores requires several assumptions, including exchangeability (no unmeasured confounding) and positivity (nonzero probability of treatment). Neither assumption is formally testable. Positivity can be addressed by ensuring that the women in the sample all have the indication for the medication (i.e., if assessing safety of antidepressants, all women in the sample should be at risk for treatment) and that no individuals with clear contraindications are included. Exchangeability is never assured; however, sensitivity analyses can yield estimates for how vulnerable an effect estimate may be to unmeasured confounding.

Strengths and Limitations: Propensity scores and other summary scores reduce many confounders to a single factor, which is especially relevant when working with a common treatment and rare outcome. They also separate the design of the study (modeling confounding) from modeling the outcome, reducing the unconscious tendency to “peek” at the outcome before the final model is decided upon. However, scenarios in which the exposure (or disease, in the case of disease risk scores) is rare, summary scores do not perform particularly well, and of course, these methods do not address unmeasured confounding. In addition, use propensity score methods in case control or
case cohort studies may produce the appearance of effect modification and/or result in residual confounding, particularly when the sample size is small.\textsuperscript{19}

\textit{I.B. Marginal Structural Models}

Marginal structural models (MSM) address time-varying exposure and confounding.\textsuperscript{20,21} Rules for confounder adjustment state we must adjust for confounders or common causes of the exposure and outcome, but should not adjust for factors on the causal pathway. In the case of time-varying exposure and confounding, we encounter a double bind: factors that are confounders in one part of the causal structure are mediators in another part (Figure S1A). For example, when studying the safety of antidepressants, we may wish to control for depression severity. However, antidepressant use in earlier pregnancy predicts depressive symptoms in later pregnancy, which will also predict subsequent antidepressant use. Standard adjustments for depression severity will always be biased in this scenario.

Central to the MSM is the inverse probability of treatment weight (IPTW). At each measurement time $t$, the investigator uses logistic regression to construct the numerator (probability of exposure) and denominator (probability of exposure, given baseline predictors and history of exposure at time $t-1$). The exposed individuals are given a weight equal to the inverse of the probability of exposure, and the unexposed are given a weight equal to the inverse of one minus the probability of exposure.\textsuperscript{22} The total weight is the product of the weights at each time point, and analyses are conducted in the weighted population, or \textit{pseudo-population}, in which individuals who are likely to be exposed (conditional on measured confounders) are given less weight, while those who are unlikely to be exposed are given more weight, resulting in balance of measured confounders within strata of exposure.

Use of marginal structural models for pregnancy medication safety studies remains rare,\textsuperscript{23,24} despite examples where timing of exposure is of great importance, and exposure is conditional on time-varying confounders, such as other medication use, changes in health status, or changes in indication or severity of disease.

\textit{Assumptions:} Under assumptions of positivity, exchangeability, and consistency, the MSM will give an unbiased estimate of the effect of the exposure on the outcome.
These assumptions are not formally testable, although assessment of the positivity assumption may include evaluation of the IPTW for extreme weights, and progressive truncation of the weights to determine whether extreme weights are highly influential.\textsuperscript{25} When important confounders are unmeasured or incompletely measured, marginal structural model methods will not provide unbiased effect estimates.

\textit{Strengths and Limitations:} The key strength of the MSM is that it allows consideration of time-varying exposure and confounding, which is highly relevant in pregnancy research due to the changes in fetal vulnerability through the course of pregnancy, and the tendency of women to change their medication use during pregnancy, even for chronic medications.\textsuperscript{26} However, when the treatment-covariate association is very strong, resulting in extreme IPTWs, MSMs can produce very wide confidence intervals which fail to include the true effect, especially for smaller sample sizes.\textsuperscript{25}

\textbf{II. Methods for incomplete confounder data}

Important confounders may be unmeasured in some databases, and failure to adjust for these unmeasured confounders results in biased effect estimates (Figure S1B). In some situations, the confounder of interest was not measured in the original dataset, but was measured in a sample that is a reasonable stand-in for the original sample. In this scenario, it is still possible to adjust for the confounder, even if the outcome has not been measured in this sample, using propensity score calibration.\textsuperscript{27-29} Propensity score calibration is a method based on regression calibration\textsuperscript{30} that offers an additional advantage over other methods of calibration, which examine residual confounding by a single variable,\textsuperscript{31} because it allows for adjustment of the joint effects of multiple confounders. For example, in a study of triptan safety, we used a cross-sectional study to jointly adjust estimates for migraine severity and type.\textsuperscript{32}

In this method, two propensity scores must be calculated: the error-prone propensity score (estimated in both the main and validation studies, including only the confounders available in the main study) and the gold-standard propensity score (estimated in the validation study, including all confounders). The outcome model is fitted using the difference between the error-prone and gold standard propensity scores to calibrate effect estimates.
Assumptions: In addition to the assumptions of propensity score models, outlined previously, propensity score calibration also assumes that the validation sample (whether internal or external) is a valid stand-in for the main sample and that the measurement error model is correctly specified.\textsuperscript{27,28} Propensity score calibration depends on an assumption of surrogacy, which means that the error-prone propensity is an adequate surrogate for the gold-standard propensity score.\textsuperscript{33} If the outcome is not measured in the validation study, the surrogacy assumption is not formally testable, although it has been shown that violations of surrogacy occur when the direction of confounding differs between the main and validation studies,\textsuperscript{27} and that bias arising from violations of surrogacy can be predicted.\textsuperscript{33}

Other methods exist for addressing partially-missing confounder data, including weighting by the inverse probability of missingness, as well as standard imputation techniques, and a comparison of these methods with propensity score calibration showed little material differences in bias reduction.\textsuperscript{34}

Strengths and Limitations: The main strength of propensity score calibration allows for adjustment for additional confounders, if they were not measured completely in the original study, even if outcome data are not available, and they have a strong theoretical base in the regression calibration literature. However, calibration methods fail when unmeasured confounding is strong, and violations of the surrogacy assumption may result in increased bias.

III. Methods for unmeasured confounding

Information on confounders may be too difficult to adequately measure (e.g., family environment or parenting style) or too costly (e.g. deep sequencing genetic data for large samples). The methods discussed below exploit aspects of observational data to control for measured and unmeasured confounders. Because siblings share substantial hereditary and environmental risk factors, comparing the outcomes of two siblings with discordant in utero exposure experiences is fairer than comparing two unrelated individuals.

III.A. Sibling Comparison Designs
If the unmeasured confounders can reasonably be thought of as shared between siblings (see Figure S1C for illustration), then studies examining sibling groups where medication exposure varies within sibling groups allows researchers to remove bias due to shared confounders.\textsuperscript{35–37} If, for example, we believe that any observed differences in autism risk between children with and without prenatal exposure to antidepressants is due to inherited genetic risk from the mother, then comparing the risk of autism between pairs of siblings with discordant prenatal exposure should be less biased than comparing autism risk between unrelated exposed and unexposed groups.

There has been substantial uptake of sibling study designs in the pregnancy medication safety literature in recent years, particularly in studies examining the safety of antidepressants, where the main concern is separating the underlying genetic and familial components of depression from exposure to antidepressant medications.\textsuperscript{38,39}

Assumptions: Use of sibling designs is most appropriate when confounders that are shared between siblings are more important than those that are unshared,\textsuperscript{36} and that there are no carryover effects between siblings.\textsuperscript{40}

Strengths and Limitations: Sibling designs have the advantage of controlling confounding, both measured and unmeasured, that is shared between siblings. However, failing to control for important unshared confounders between siblings results in increased bias; sibling studies are also more vulnerable to bias from measurement error than non-sibling studies.\textsuperscript{36}

III.B. Instrumental Variables

Instrumental variable (IV) methods \textsuperscript{41,42} require identifying a variable or variables whose effect on the outcome occurs only through the exposure: an example of a perfect instrument is a coin toss assigning an individual to exposure or non-exposure, while commonly used (and less strong) instruments include provider prescription preference and distance from health care services. One example of instrumental variable methods used in pregnancy medication research is a study of antidepressant (AD) efficacy during pregnancy using provider preference, calendar time as a function of recent FDA recommendations, and geographic differences in AD use as instruments; however, these instruments were only weakly associated with the treatment, which may have contributed to the equivocal findings from this study.\textsuperscript{43} IV studies are most often conducted using a
For Review Only

[Page 12 of 37]

Pharmacoepidemiology and Drug Safety

two stage least squares methods, where in the first stage, the instrument or instruments are used as explanatory variables in a model predicting the exposure, and the predicted values from this first stage are used as predictors in a model of the outcome. In general, identifying a strong instrument that meets all of the necessary assumptions is extremely challenging, which has likely contributed to the slower adoption of this method.

Mendelian randomization, which uses a genetic marker as an instrument, is a subtype of instrumental variable analysis;\textsuperscript{44} while Mendelian randomization has not yet been used in pregnancy medication studies, studies estimating the effect of alcohol use during pregnancy on later neurocognitive outcomes have used the genetic variants encoding alcohol dehydrogenase, an enzyme that metabolizes alcohol, with some success.\textsuperscript{45}

\textit{Assumptions:} Instrumental variable analyses allow for an unbiased effect estimate under certain strict assumptions: (i) the instrument has a causal effect on the exposure of interest, (ii) the instrument effects the outcome only through the exposure of interest, not through any other pathways, (iii) there are no common causes or confounders of the instrument-outcome pathway (Figure S1D).

\textit{Strengths and Limitations:} Instrumental variable analyses address bias from measured and unmeasured confounders, and so instruments that meet all the assumptions will mimic the results from a randomized trial. However, estimates are highly sensitive to violations of untestable assumptions, and violations may produce bias amplification; for example, if there is unmeasured confounding of the instrument and outcome.\textsuperscript{41}

Figure 2 guides readers through the process of selecting a method or methods, based on categorization of confounder data according to availability and timing. The most important first step is to draw a DAG or DAGs that represent the proposed causal mechanism, without regard to availability of data on confounders: if a confounder is important, it should be included in the DAG, even if the study did not collect data on it. The next step is to determine which confounders are available in your study, and whether the data support the appropriate analytic method. For example, if your DAG shows that in theory, medication use and confounders vary over time, but your data shows no such variation, a MSM approach should not be used. In addition to addressing measured
confounders, methods for dealing with unmeasured or incompletely measured confounders should be considered, and again, the researcher must determine whether the data available to her can support the analytic method: if the data cannot identify siblings, or if the assumptions of instrumental variables are clearly not met, these methods cannot be used. Finally, quantitative bias analysis should be employed as a sensitivity analysis. Most importantly, we urge researchers to consider potential sources of confounding regardless of whether they were measured in the data, and to choose the methods most suited to the data they have available: Figure 2 suggests a systematic way of approaching this process.

A reference to selected software for the methods discussed in this paper is included as part of the supplemental material.

Discussion

Studies of medication use during pregnancy depend upon observational data to answer critical questions of safety and efficacy. More traditional methods for confounding control, such as stratification, restriction, matching, and adjustment have been described in great detail elsewhere, and because of this we have not discussed them here. These older methods have their place in observational research, but as our understanding of the complexities of bias has progressed, so has our understanding of the limitations of these methods. The advanced methods described in this paper were developed to address specific confounding problem, and are necessary to reduce bias, and ultimately to produce the best information possible to health care providers and pregnant women. Using these methods can produce substantially different results from traditional methods, such as when we compare the cohort and sibling studies of antidepressant safety, the regression-adjusted to the marginal structural model estimates for triptan safety, propensity calibrated estimates to standard propensity score methods, or including standard regression, propensity score methods, sibling controls, and negative paternal controls in one study.

Aside from a few examples, these methods have seen slow uptake in the pregnancy medication research literature. This is almost certainly due to a sense of caution about methods that can seem opaque, especially upon first encounter with the seminal methods paper describing the technique. It is true that caution is necessary when
applying novel methods. However, it is also true that the more standard regression methods require similar assumptions to the more advanced methods discussed in this paper. Perhaps it is true that familiarity with standard methods had bred, if not contempt, then certainly comfort. If readers find that their research question fits well with one of the scenarios described in this paper, we suggest approaching the problem by tackling the citations given for the technique. The techniques we describe in this paper have their roots in standard regression techniques and can be implemented with standard software.

While this paper focuses on bias due to confounding, other sources of bias such as exposure and/or outcome misclassification and selection bias, as well as seasonal effects, can also distort effect estimates. This paper is not intended to be an exhaustive discussion of all possible methods for confounding control. New techniques are being developed all the time, and many of these, such as g-estimation and targeted maximum likelihood estimation, have not yet been implemented in the pregnancy medication literature. Quantitative bias analysis can help researchers account for bias due to systematic errors in their data, and these methods are easily implemented with standard software. Further, the methods discussed herein are not mutually exclusive, and can be used in combination with each other: combining propensity scores with instrumental variables or marginal structural models with quantitative bias analysis gives far more information about the probable range of effect estimates than any of these methods alone.

Observational pharmacoepidemiology studies are vital to our understanding of the safety of medications in pregnancy, but as they are based on observational data, great care must be taken in the analysis and interpretation of data to minimize confounding and bias. In all pharmacoepidemiological studies sources of bias should be acknowledged and discussed, and preferably quantified by performing sensitivity analysis of estimates under an array of assumptions about possible bias directions and magnitudes.
References


12. Schneeweiss S, Rassen J a, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-

http://mc.manuscriptcentral.com/pds


46. Brandlistuen RE, Ystrom E, Eberhard-Gran M, Nulman I, Koren G, Nordeng H.


FIGURE LEGENDS

Figure 1. Conceptual model for the effect of prenatal SSRI exposure on attention deficit/hyperactivity disorder (ADHD), including a set of important confounders (depression severity, concomitant medication use, genetics), a potential mediator (gestational age), a collider (live birth), and factors related only to the exposure (pre-pregnancy SSRI use) or the outcome (child gender).

Figure 2. Choosing methods for confounding control
Table 1. Examples of application of advanced confounding control methods in the pregnancy medication safety literature.

<table>
<thead>
<tr>
<th>Medication &amp; Study Reference</th>
<th>Outcome</th>
<th>Confounder(s)</th>
<th>Confounding Problem(s)</th>
<th>Method(s) Employed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ondansetron (Pasternak, Svanström, &amp; Hviid, 2013)</td>
<td>Malformations</td>
<td>Nausea/vomiting; maternal characteristics, comorbidities, other medications, pregnancy history.</td>
<td></td>
<td>Propensity scores/summary scores, Marginal structural models</td>
</tr>
<tr>
<td>Lithium (Patorno et al., 2017)</td>
<td>Cardiac Malformations</td>
<td>Maternal comorbidities, other medications, maternal characteristics.</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Statins (Bateman et al., 2015)</td>
<td>Malformations</td>
<td>Maternal characteristics, obstetric and medical conditions, other medications.</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Triptans (Wood, Lapane, et al., 2015)</td>
<td>Neurodevelopment</td>
<td>Other medications (time-varying), maternal characteristics; migraine severity.</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Iron supplementation (Bodnar, Davidian, Siega-Riz, &amp; Tsiatis, 2004)</td>
<td>Anemia</td>
<td>Maternal baseline characteristics; gastric symptoms; serum ferritin and hemoglobin concentration.</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Triptans (Wood, Frazier, Nordeng, &amp; Lapane, 2015)</td>
<td>Neurodevelopment</td>
<td>Other medications, maternal characteristics; migraine severity, attitudes about medication use.</td>
<td>x x x</td>
<td></td>
</tr>
<tr>
<td>SSRI (Nezvalová-Henriksen et al., 2016; Viktorin et al., 2016)</td>
<td>Gestational age, birth weight</td>
<td>Family factors, maternal depression; illnesses, other medications.</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Anti-epileptic drugs (Bech et al., 2014)</td>
<td>Spontaneous abortion</td>
<td>Severity of maternal epilepsy; maternal characteristics, environmental exposures, comorbidities.</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SSRI (Swanson et al., 2015)</td>
<td>Maternal depression relapse</td>
<td>Maternal depression severity; comorbidities, other medications, maternal characteristics, proxies for severity.</td>
<td>x x</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Conceptual model for the effect of prenatal SSRI exposure on attention deficit/hyperactivity disorder (ADHD), including a set of important confounders (depression severity, concomitant medication use, genetics), a potential mediator (gestational age), a collider (live birth), and factors related only to the exposure (pre-pregnancy SSRI use) or the outcome (child gender).
Figure 2. Choosing methods for confounding control

254x190mm (72 x 72 DPI)
Figure S1. Directed acyclic graphs (DAGs) for (a) Time varying confounding: time-varying exposure $A$, outcome $Y$, baseline confounders $C$ and time-varying confounders $TVC$ at times 0, 1, and 2; (b) Unmeasured confounding: exposure $A$, outcome $Y$, and measured $C$ and unmeasured $U$ confounders; (c) Sibling study design, for siblings (1 and 2), with exposure $A$, outcome $Y$, and confounders $C$ of $AY$, and shared unmeasured factors which cause $C$, $A$, and $Y$; (d) Instrumental variable (IV) which affects the outcome $Y$ only through the exposure $A$ and therefore controls both measured confounders $C$ and unmeasured confounders $U$. 
Box 1. Illustration of confounding control using propensity matching and weighting.

Women who use a medication during pregnancy ("treated") often have substantially different characteristics than women who do not use a medication ("untreated"); in this illustration, different characteristics are shown through shading and patterns in the figures. Part A shows that the treated and untreated groups are confounded by several participant characteristics.

For example:

Participant 1 has severe depression and drug abuse problems.

Participant 2 has mild anxiety and a history of smoking.

When we use propensity score matching to control for measured confounders, we may be matching two participants who are not similar in terms of specific confounders, but whose propensity for treatment is the same.

When we compare the treated and untreated groups in Part B, we see that the individual participants in the treated and untreated groups look quite different. However, the distribution of characteristics in each group is the same.

It is also important to note that in the matched sample, a large number of the untreated group has been discarded.

Another use of the propensity score is to use the propensity score to weight the study sample. Weights in the sample vary between zero and infinity, and care must be taken to avoid bias from extreme weights.

Weighted analyses are conducted in a pseudo-population. In this pseudo-population, participants with rare combinations of confounders are up-weighted (e.g., they count for more than one participant in the analysis) and participants with common combinations of confounders are down-weighted (e.g., they count for less than one participant in the analysis). This results in balanced confounders within each stratum of treatment.

In this illustration (Part C), the participant characteristics are balanced within each stratum of treatment, and we have used lighter drawings to illustrate the presence of a pseudo-population.
<table>
<thead>
<tr>
<th>Method</th>
<th>Software</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propensity Scores</td>
<td>SAS</td>
<td>Macros “psmatch_multi”, “cem” available for public use</td>
</tr>
<tr>
<td></td>
<td>Stata</td>
<td>tefects command available in v13 and later; user-written</td>
</tr>
<tr>
<td></td>
<td></td>
<td>commands “pscore”, “match”, “cem”, “psmatch2” for earlier</td>
</tr>
<tr>
<td></td>
<td></td>
<td>versions</td>
</tr>
<tr>
<td>Notes:</td>
<td></td>
<td>all major software</td>
</tr>
<tr>
<td></td>
<td></td>
<td>packages can output</td>
</tr>
<tr>
<td></td>
<td></td>
<td>predicted probabilities after</td>
</tr>
<tr>
<td></td>
<td></td>
<td>logistic regression to give</td>
</tr>
<tr>
<td></td>
<td></td>
<td>propensity score, which can</td>
</tr>
<tr>
<td></td>
<td></td>
<td>be used to calculate weights</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and strata. Additional</td>
</tr>
<tr>
<td></td>
<td></td>
<td>software can assist with</td>
</tr>
<tr>
<td></td>
<td></td>
<td>matching,</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Packages “twang”, “cem”, “optmatch”, “matchit” and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“matching” available for download</td>
</tr>
<tr>
<td>Marginal Structural Models</td>
<td>SAS</td>
<td>After creating joint IPTW in data step, use “weight”</td>
</tr>
<tr>
<td></td>
<td>Stata</td>
<td>option in standard outcome analyses.</td>
</tr>
<tr>
<td></td>
<td>SPSS</td>
<td>No formal support; “PS Matching” add-on calls</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>“matchit” package from R.SPSS only supports</td>
</tr>
<tr>
<td></td>
<td></td>
<td>frequency weights except through the complex</td>
</tr>
<tr>
<td></td>
<td></td>
<td>survey sampling add-on! Software may round</td>
</tr>
<tr>
<td></td>
<td></td>
<td>propensity weight to the nearest whole number</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and use it as a frequency weight without returning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>an error message.</td>
</tr>
<tr>
<td>Notes:</td>
<td></td>
<td>weights estimated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>from pooled logistic regression, similarly to PS methods above.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Notes above regarding weights in SPSS also apply here.</td>
</tr>
<tr>
<td>Propensity Calibration</td>
<td>SAS</td>
<td>Carry out through regression calibration; “Blinplus”</td>
</tr>
<tr>
<td></td>
<td>Stata</td>
<td>macro available for download</td>
</tr>
<tr>
<td></td>
<td>SPSS</td>
<td>Not available.</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>“iWeigReg” package after propensity score estimation.</td>
</tr>
<tr>
<td>Notes:</td>
<td></td>
<td>technique developed using SAS macros. Other software supports</td>
</tr>
<tr>
<td></td>
<td></td>
<td>regression calibration (noted at right) but has not been testing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in the unmeasured confounding application.</td>
</tr>
<tr>
<td>Instrumental Variable(s)</td>
<td>SAS</td>
<td>PROC SYSLIN, PROC MODEL, PROC CALIS</td>
</tr>
<tr>
<td></td>
<td>Stata</td>
<td>“IVRegress” included from v13 on; user written</td>
</tr>
<tr>
<td></td>
<td></td>
<td>package “ivreg2” fits 2sls models; other packages</td>
</tr>
<tr>
<td></td>
<td>SPSS</td>
<td>available for extended applications.</td>
</tr>
<tr>
<td>Notes:</td>
<td></td>
<td>multiple approaches to IV analyses. Two stage least squares</td>
</tr>
<tr>
<td></td>
<td></td>
<td>approaches are</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROC MIXED and PROC NLMIXED</td>
</tr>
<tr>
<td>Sibling Comparison</td>
<td>SAS</td>
<td>Xtreg command</td>
</tr>
<tr>
<td></td>
<td>Stata</td>
<td>User-written command</td>
</tr>
<tr>
<td></td>
<td>SPSS</td>
<td>“rcal” performs regression calibration</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>“iWeigReg” package after propensity score estimation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“lme” and “nlme”</td>
</tr>
</tbody>
</table>
most appropriate for linear outcomes; non-linear outcomes require special consideration, particularly with respect to standard errors.
CONFLICT OF INTEREST DISCLOSURE

The Editors of Pharmacoepidemiology and Drug Safety recognize that most studies in pharmacoepidemiology cost money and thus pose a potential conflict of interest. As a conflict of interest may affect the assessment or judgment of an author, we ask that all authors (not just the Corresponding Author) complete the following form.

For Co-authors: Please complete questions 4-10. Completed forms should be saved, and emailed as an attachment to the Corresponding Author.

For Corresponding Authors: Please complete all questions. It is the responsibility of the Corresponding Author to submit completed forms on behalf of all co-authors via Manuscript Central at the point of manuscript submission.

### Corresponding author only (Co-authors go to Question 4):

**POTENTIAL STUDY INTERPRETATION CONFLICTS**

1. Some or all of the data that were used in this study were provided by a company with a vested interest in the product being studied. No

2. The sponsor of this project had the right of commenting but the authors retained the right to accept or reject comments or suggestions. n/a

3. The sponsor of this project had the right of final editing and/or approval of the manuscript submitted. n/a

**Corresponding author and Co-authors:**

**POTENTIAL FINANCIAL CONFLICTS**

4. I, my spouse, or one of my dependent children is an employee of a company whose product is being studied. No

5. I, my spouse, or one of my dependent children has significant equity interest (>USD 10,000) in the company that owns the product being studied. No
6. In the past three years I have:
   - been paid as a consultant (or in a similar capacity) by a company with a vested interest in the product being studied, on issues related to the product being studied; No
   - been paid as a consultant (or in a similar capacity by a company with a vested interest in the product being studied, on issues unrelated to the product being studied; No
   - received research or educational support from a company with a vested interest in the product(s) being studied. No

7. A company whose product is being studied has provided funding to support the work on this project. No

If you have answered YES to any of the above questions, or if you have additional personal, commercial or academic conflicts of interest, please draft a statement to publish with the article. e.g., AB has been reimbursed by Safe Drug Ltd. for international conference attendance.

8. Manuscript title (first six words are sufficient)
   Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control

9. Author's full name (a separate form must be submitted for each author)
   Mollie Wood

10. In checking this box, I confirm I have completed this form to the best of my knowledge. ☒

This form is available online by clicking here
CONFLICT OF INTEREST DISCLOSURE

The Editors of Pharmacoepidemiology and Drug Safety recognize that most studies in pharmacoepidemiology cost money and thus pose a potential conflict of interest. As a conflict of interest may affect the assessment or judgment of an author, we ask that all authors (not just the Corresponding Author) complete the following form.

For Co-authors: Please complete questions 4-10. Completed forms should be saved, and emailed as an attachment to the Corresponding Author.

For Corresponding Authors: Please complete all questions. It is the responsibility of the Corresponding Author to submit completed forms on behalf of all co-authors via Manuscript Central at the point of manuscript submission.

Corresponding author only (Co-authors go to Question 4):
POTENTIAL STUDY INTERPRETATION CONFLICTS

1. Some or all of the data that were used in this study were provided by a company with a vested interest in the product being studied. No

2. The sponsor of this project had the right of commenting but the authors retained the right to accept or reject comments or suggestions. n/a

3. The sponsor of this project had the right of final editing and/or approval of the manuscript submitted. n/a

Corresponding author and Co-authors:
POTENTIAL FINANCIAL CONFLICTS

4. I, my spouse, or one of my dependent children is an employee of a company whose product is being studied. No

5. I, my spouse, or one of my dependent children has significant equity interest (>USD 10,000) in the company that owns the product being studied. No
6. In the past three years I have:

- been paid as a consultant (or in a similar capacity) by a company with a vested interest in the product being studied, on issues related to the product being studied; No
- been paid as a consultant (or in a similar capacity by a company with a vested interest in the product being studies, on issues unrelated to the product being studied; No
- received research or educational support from a company with a vested interest in the product(s) being studied. No

7. A company whose product is being studied has provided funding to support the work on this project. No

If you have answered YES to any of the above questions, or if you have additional personal, commercial or academic conflicts of interest, please draft a statement to publish with the article. e.g., AB has been reimbursed by Safe Drug Ltd. for international conference attendance.

8. Manuscript title (first six words are sufficient)

Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control

9. Author’s full name (a separate form must be submitted for each author)

Marleen M.H.J. van Gelder

10. In checking this box, I confirm I have completed this form to the best of my knowledge. ☑️
CONFLICT OF INTEREST DISCLOSURE

The Editors of *Pharmacoepidemiology and Drug Safety* recognize that most studies in pharmacoepidemiology cost money and thus pose a potential conflict of interest. As a conflict of interest may affect the assessment or judgment of an author, we ask that all authors (not just the Corresponding Author) complete the following form.

For Co-authors: Please complete questions 4-10. Completed forms should be saved, and emailed as an attachment to the Corresponding Author.

For Corresponding Authors: Please complete all questions. It is the responsibility of the Corresponding Author to submit completed forms on behalf of all co-authors via Manuscript Central at the point of manuscript submission.

<table>
<thead>
<tr>
<th><strong>Corresponding author only (Co-authors go to Question 4):</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POTENTIAL STUDY INTERPRETATION CONFLICTS</strong></td>
</tr>
<tr>
<td>1. Some or all of the data that were used in this study were provided by a company with a vested interest in the product being studied.</td>
</tr>
<tr>
<td>2. The sponsor of this project had the right of commenting but the authors retained the right to accept or reject comments or suggestions.</td>
</tr>
<tr>
<td>3. The sponsor of this project had the right of final editing and/or approval of the manuscript submitted.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Corresponding author and Co-authors:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POTENTIAL FINANCIAL CONFLICTS</strong></td>
</tr>
<tr>
<td>4. I, my spouse, or one of my dependent children is an employee of a company whose product is being studied.</td>
</tr>
<tr>
<td>5. I, my spouse, or one of my dependent children has significant equity interest (&gt;USD 10,000) in the company that owns the product being studied.</td>
</tr>
</tbody>
</table>
6. In the past three years I have:

- been paid as a consultant (or in a similar capacity) by a company with a vested interest in the product being studied, on issues related to the product being studied; No
- been paid as a consultant (or in a similar capacity by a company with a vested interest in the product being studied, on issues unrelated to the product being studied; No
- received research or educational support from a company with a vested interest in the product(s) being studied. No

7. A company whose product is being studied has provided funding to support the work on this project. No

If you have answered YES to any of the above questions, or if you have additional personal, commercial or academic conflicts of interest, please draft a statement to publish with the article. e.g., AB has been reimbursed by Safe Drug Ltd. for international conference attendance.

N/A

8. Manuscript title (first six words are sufficient)

Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control

9. Author’s full name (a separate form must be submitted for each author)

Dheeraj Rai

10. In checking this box, I confirm I have completed this form to the best of my knowledge.

This form is available online by clicking here
CONFLICT OF INTEREST DISCLOSURE

The Editors of Pharmacoepidemiology and Drug Safety recognize that most studies in
pharmacoepidemiology cost money and thus pose a potential conflict of interest. As a conflict of interest
may affect the assessment or judgment of an author, we ask that all authors (not just the Corresponding
Author) complete the following form.

For Co-authors: Please complete questions 4-10. Completed forms should be saved, and emailed as an
attachment to the Corresponding Author.

For Corresponding Authors: Please complete all questions. It is the responsibility of the Corresponding
Author to submit completed forms on behalf of all co-authors via Manuscript Central at the point of
manuscript submission.

Corresponding author only (Co-authors go to Question 4):

POTENTIAL STUDY INTERPRETATION CONFLICTS

1. Some or all of the data that were used in this study were provided by a company with a vested
interest in the product being studied. [No]

2. The sponsor of this project had the right of commenting but the authors retained the right to
accept or reject comments or suggestions. [n/a]

3. The sponsor of this project had the right of final editing and/or approval of the manuscript
submitted. [n/a]

Corresponding author and Co-authors:

POTENTIAL FINANCIAL CONFLICTS

4. I, my spouse, or one of my dependent children is an employee of a company whose product is
being studied. [n/a]

5. I, my spouse, or one of my dependent children has significant equity interest (>USD 10,000) in the
company that owns the product being studied. [n/a]
6. In the past three years I have:

- been paid as a consultant (or in a similar capacity) by a company with a vested interest in the product being studied, on issues related to the product being studied: n/a
- been paid as a consultant (or in a similar capacity by a company with a vested interest in the product being studied, on issues unrelated to the product being studied: n/a
- received research or educational support from a company with a vested interest in the product(s) being studied.

7. A company whose product is being studied has provided funding to support the work on this project.

If you have answered YES to any of the above questions, or if you have additional personal, commercial or academic conflicts of interest, please draft a statement to publish with the article. e.g. AB has been reimbursed by Safe Drug Ltd. for international conference attendance.

8. Manuscript title (first six words are sufficient)

Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control

9. Author's full name (a separate form must be submitted for each author)

Kate Lapanne

10. In checking this box, I confirm I have completed this form to the best of my knowledge. ☑

This form is available online by clicking here.

September 2016

http://mc.manuscriptcentral.com/pds
CONFLICT OF INTEREST DISCLOSURE

The Editors of Pharmacoepidemiology and Drug Safety recognize that most studies in pharmacoepidemiology cost money and thus pose a potential conflict of interest. As a conflict of interest may affect the assessment or judgment of an author, we ask that all authors (not just the Corresponding Author) complete the following form.

For Co-authors: Please complete questions 4-10. Completed forms should be saved, and emailed as an attachment to the Corresponding Author.

For Corresponding Authors: Please complete all questions. It is the responsibility of the Corresponding Author to submit completed forms on behalf of all co-authors via Manuscript Central at the point of manuscript submission.

---

**Corresponding author only (Co-authors go to Question 4):**

**POTENTIAL STUDY INTERPRETATION CONFLICTS**

1. Some or all of the data that were used in this study were provided by a company with a vested interest in the product being studied. No

2. The sponsor of this project had the right of commenting but the authors retained the right to accept or reject comments or suggestions. n/a

3. The sponsor of this project had the right of final editing and/or approval of the manuscript submitted. n/a

---

**Corresponding author and Co-authors:**

**POTENTIAL FINANCIAL CONFLICTS**

4. I, my spouse, or one of my dependent children is an employee of a company whose product is being studied. n/a

5. I, my spouse, or one of my dependent children has significant equity interest (>USD 10,000) in the company that owns the product being studied. n/a

---

September 2016

http://mc.manuscriptcentral.com/pds
6. In the past three years I have:
   
   - been paid as a consultant (or in a similar capacity) by a company with a vested interest in the product being studied, on issues related to the product being studied; n/a
   
   - been paid as a consultant (or in a similar capacity by a company with a vested interest in the product being studied, on issues unrelated to the product being studied; n/a
   
   - received research or educational support from a company with a vested interest in the product(s) being studied. n/a

7. A company whose product is being studied has provided funding to support the work on this project. n/a

If you have answered YES to any of the above questions, or if you have additional personal, commercial or academic conflicts of interest, please draft a statement to publish with the article. e.g., AB has been reimbursed by Safe Drug Ltd. for international conference attendance.

8. Manuscript title (first six words are sufficient)
   
   Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control

9. Author’s full name (a separate form must be submitted for each author)
   
   Hedvig Nordeng

10. In checking this box, I confirm I have completed this form to the best of my knowledge. ☒

This form is available online by clicking here.