Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2019 July ; 28(7): 906–922. doi:10.1002/pds.4789.

Use of real-world evidence from healthcare utilization data to evaluate drug safety during pregnancy

Krista F. Huybrechts¹, Brian T. Bateman^{1,2}, Sonia Hernández-Díaz³

¹Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

²Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

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

Abstract

Purpose: Because preapproval clinical trials typically exclude pregnant women, the evidence on drug safety during pregnancy required to inform drug labeling must come from postapproval controlled observational studies. Common designs have included pregnancy registries and case-control studies. Recently, pregnancy cohorts nested within healthcare utilization databases are increasingly being used. Despite clear advantages, these databases share some important limitations that may threaten the validity of studies emerging from them.

Methods: This paper describes the distinctive methodological aspects of conducting drug safety studies in healthcare utilization databases with special emphasis on design and analytic approaches to minimize biases.

Results: We describe considerations for study design, cohort definition, and follow-up. We then address issues related to exposure ascertainment based on prescription fills, including the importance of the etiologically relevant window and of properly accounting for preterm births. This is followed by a discussion of advantages and challenges when ascertaining maternal and infant outcomes based on secondary data. We then explore useful approaches to address confounding within the context of pregnancy research and of the potential for selection bias when restricting the cohort to live births. Finally, we consider issues related to external validity and statistical significance. The examples are mainly drawn from a pregnancy cohort nested in the Medicaid Analytic Extract.

Conclusions: The approaches presented provide guidance regarding the important methodological considerations that need to be attended to in order to generate valid, minimally

Correspondence K. F. Huybrechts, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, 1620 Tremont Street, Suite 3030, Boston, MA. khuybrechts@bwh.harvard.edu.

ETHICS STATEMENT

The authors state that no ethical approval was needed.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

biased risk when using large healthcare utilization databases for drug safety surveillance in pregnancy.

Keywords

drug safety; guidance; healthcare utilization; pharmacoepidemiology; pregnancy

1 | INTRODUCTION

Preapproval clinical trials typically exclude pregnant women.¹ Hence, the evidence on pregnancy safety required for drug labels must come from postapproval controlled observational studies. Regulatory agencies are increasingly recommending to complement pregnancy registries^{2,3} and case-control studies⁴ with pregnancy cohorts nested within healthcare utilization databases such as national registries (eg, Nordic registers), electronic medical records (eg, Clinical Practice Research Datalink), and insurance claims (eg, Truven MarketScan Research Database). These databases provide prospectively collected information for large populations, enable the study of multiple drugs and multiple outcomes, and are more efficient in terms of time and cost than primary data collection. Moreover, they reflect routine care and include populations often excluded from volunteer-based studies.

Despite many advantages, studies emerging from secondary analysis of data collected for clinical documentation or administrative purposes share some limitations that may threaten their validity. In addition, each database might present specific constraints (eg, lack of information from specialist or outpatient encounters, gestational age, or use of nonprescription drugs). If these limitations are not well understood and addressed, studies based on these data sources can misinform risk-benefit decisions for drugs. The main goal of this review is to provide guidance for researchers regarding how to use these data in a valid manner to study drug safety in pregnancy.

2 | DATA SOURCE FOR EXAMPLES USED IN THIS REVIEW

To illustrate the approaches, we draw examples from the Medicaid Analytic Extract (MAX) pregnancy cohort,⁵ but issues equally apply to most healthcare utilization databases. Medicaid data provide a unique opportunity to examine drug safety questions in a national cohort of pregnant women in the United States, where Medicaid covers almost half of the deliveries.⁶ The MAX pregnancy cohort includes a preponderance of young women, racial and ethnic minorities, and disabled persons. MAX records beneficiary demographic and enrollment information, all diagnoses and procedures associated with inpatient hospitalizations and outpatient visits. It also contains claims for all filled outpatient medication prescriptions. Diagnoses are coded with the clinical modification of the International Classification of Diseases (ICD)-9 system (and ICD-10 after 2015) and procedures with the Current Procedural Terminology (CPT)-4. The pharmacy file records claims for each prescription fill including the date of dispensing, the drug, the quantity, the days the supply is anticipated to last, and the dose.

The MAX pregnancy cohort offers many strengths, including its large population-based cohort (about 1.8 million pregnancy for 2000-2013 from 46 US states and Washington, DC)

which reduces random error and increases generalizability, and linkage to medical records for validation studies and collection of covariates not well captured in claims data. However, it also presents the challenges shared by any data source of this type and steps needed to be taken to ensure internal validity. We have employed this data source to study the safety of drugs during pregnancy with respect to a wide variety of important fetal and maternal outcomes.⁷⁻¹⁶

3 | STUDY DESIGN AND STUDY POPULATION

The study design must be consistent with the scientific question being addressed. For drug safety in pregnancy, we want to know “Is this treatment safe (relative to its benefits) for a patient who is pregnant or considering a pregnancy?” To respond to this question in the absence of randomization, the design needs to ensure that women in the exposed and reference groups only differ in their drug exposure, and not in terms of risk factors for the outcome. We will focus on the cohort design even though other approaches can be applied to healthcare utilization databases including nested case-control and case-only designs.¹⁷

Conceptually, the study population will be defined through specific inclusion and exclusion criteria so that it closely resembles the women eligible for treatment in actual clinical practice. Operationally, studies first need to identify pregnancies using pregnancy-related diagnostic and procedure codes from healthcare utilization claims and then impose eligibility criteria to ensure complete information: (1) continuous enrollment in the insurance for a given baseline period, (2) full coverage including prescription benefits, (3) appropriate enrollment type (eg, fee for service, capitated plans only if they do not under-report encounter claims), and (4) linkage to infants for pregnancies ending in a live birth.^{18,19} In MAX, we used a deterministic linkage algorithm to link mother-infant data files based on state, Medicaid case number (which identifies family units), and date of delivery.⁵ Then, diagnosis and procedure codes and medications can be used to apply inclusion criteria (eg, pregnant women with multiple sclerosis).

The specific maternal and infant eligibility requirements will depend on the research question at hand. Requiring maternal enrollment prior to the date of last menstrual period (LMP) allows for identification of prescriptions filled prior to the LMP and provides accurate ascertainment of comorbidities that predate pregnancy (ie, baseline characteristics). The requirement of continuous enrollment throughout pregnancy allows for complete follow-up and complete ascertainment of drug exposures, diagnoses, and procedures. Requiring enrollment of infants for at least 90 days following birth allows ascertainment of most major congenital birth defects. Most cohort studies would just censor subjects as they are lost to follow-up, but in pregnancy, we often require a minimum follow-up because identification of pregnancy itself or ascertainment of outcomes might only occur at delivery.

These requirements are crucial to maximize the validity of the study but come at the cost of sample size and of potentially selecting a study population different from the source population. For most data sources, the sample size declines as increasingly longer enrollment requirements are applied to the mother and infant (Figure 1). However, we compared key clinical and demographic characteristics for pregnant women included and

excluded from the study cohort because of various eligibility requirements and observed minimal differences, except for a higher proportion of women eligible for Medicaid because of disability since women below a poverty threshold qualify only when pregnant.⁹

One distinctive characteristic of pregnancy studies is that the population can be defined at the level of women, pregnancies, or fetuses. A woman may have more than one pregnancy during the study period, and she may carry more than one fetus in the same pregnancy. Clustered analyses are often used in these situations to account for the correlation within mother and within pregnancy.²⁰ There are several ways to deal with multiple gestations from an analytical point of view, but it may be difficult to distinguish outcomes from different fetuses in administrative data. Thus, when reporting risks, it may be preferable to use pregnancies as the unit of analysis (ie, if one of the fetuses is affected by the outcome of interest, the pregnancy is considered affected). For some drugs (eg, infertility treatments), multiples are an outcome of interest or an important mediator since twinning is associated with an increased risk of many pregnancy outcomes.²¹ Multiples, like sex of the infant, are rarely relevant confounders since they do not affect the indication for most maternal drugs, with some possible exceptions such as nausea and vomiting or hypertension. Yet, they might be effect modifiers.

Another peculiarity of pregnancy studies is that drug use during pregnancy reflects by definition new use for the fetus and the pregnancy.²² Hence, the potential selection bias of including prevalent users in the study population, which is a concern in many pharmacoepidemiological studies, is only a concern when the pregnancy outcome of interest shares common causes with the pre-pregnancy treatment decisions regarding continuation. For example, this might occur when studying the association between drugs (eg, antipsychotic medications) and the risk of gestational diabetes. Prevalent users of antipsychotics at conception may represent a selection of women not vulnerable to a metabolic syndrome induced by these drugs. Since vulnerability to metabolic syndrome is also a risk factor for gestational diabetes, the inclusion of prevalent antipsychotic users without pregestational diabetes may represent a selection bias.²³

4 | FOLLOW-UP

An ideal pregnancy cohort would include women at conception and follow them while they are at risk for the outcome (eg, end of pregnancy for preeclampsia, years beyond delivery for autism). Conception is considered an acceptable time zero for prevalent pre-pregnancy drug use since it is when the risk for fetal and pregnancy outcomes begins. However, claims databases do not contain a field for conception, LMP, or gestational age at birth; thus, this information is estimated using diagnostic and procedure codes indicative of gestational age. The first day of the LMP is estimated by subtracting the gestational age at delivery from the pregnancy end date. The LMP date can be defined using validated algorithms,^{24,25} which have been demonstrated to have both a high sensitivity (>96%) and specificity (>99%) for correctly classifying trimester-specific exposure status for medications used chronically.²⁶ The performance of these algorithms for correctly classifying gestation week-specific exposure status has not previously been established. More challenging is also the timing of pregnancy for pregnancy losses.²⁷

An important consideration in pregnancy cohorts is that the baseline risks for many outcomes change during follow-up. For example, the risk of pregnancy losses is highest during the first trimester, while the risk of preeclampsia increases towards the end of pregnancy. For some outcomes, women are only at risk later in pregnancy, and the cohort should be restricted to pregnancies maintained until then (eg, to be at risk of stillbirth, a gestation must get to 20 weeks; the origins of neonatal opioid withdrawal syndrome might start weeks before delivery). Thus, early pregnancy losses compete with the risks for late pregnancy events (eg, preeclampsia, stillbirths, neonatal withdrawal). Depending on the drug mechanism of action and carryover effects, the definition of the relevant exposure window might coincide with the initiation of the period at risk or start earlier.

Follow-up needed to ascertain some pregnancy outcomes goes beyond their onset. Despite prenatal testing, malformations are often diagnosed after birth. Longer follow-ups will identify additional congenital problems because a few structural and many developmental disorders may only manifest months or years later. For example, the frequency of major malformations varies from 2.1% when diagnosed prenatally or within the first week of life to 2.6% if infants are followed until their first birthday.^{28,29} Thus, depending on the outcomes of interest, follow-up can extend into adulthood, although end of pregnancy or childhood is a more common goal. Note that when defining a minimal follow-up time, such as “enrolled for at least 3 months after birth,” a shorter follow-up should be allowed for infants who die to avoid excluding lethal outcomes.

5 | EXPOSURE

In healthcare utilization databases, maternal exposure to drugs is derived from pharmacy dispensing records, with exposure status on any given day determined by the dispensing date and number of days supply. Correctly classifying the exposure status relevant to the outcome of interest is dependent on ascertaining when the patient took the drug and defining the etiologically relevant period of exposure.

5.1 | Using databases to ascertain exposure

Prescription databases provide several advantages in defining exposure. Pharmacists fill prescriptions as written, and reimbursement by insurers requires that accurate claims be submitted. Nonresponse and recall bias are absent from claims databases since all data recording is independent of a patient's memory or agreement to participate in a research study. They also offer the possibility of estimating the timing of exposure with precision, which may be challenging to do through a retrospective interview that occurs long after the medication was used.

However, there are also important challenges. Exposure is defined based on filled prescriptions for a medication. Just because a prescription is dispensed does not mean that the patient necessarily consumes a medication. The correspondence between drug dispensing and exposure might be better for some pharmaceuticals (eg, vaccine injection, anticonvulsants for epilepsy) than for others (eg, analgesics, asthma rescue medications). In addition to the usual issues regarding adherence that challenge pharmacoepidemiologic studies, for pregnancy studies, there is the added challenge that women may deliberately

discontinue a medication prior to exhausting its supply either in anticipation of pregnancy or upon learning of conception. These issues have the potential to lead to exposure misclassification (ie, false positives), which can bias risk estimates towards those in the unexposed group. On the other hand, old prescriptions can be used during pregnancy (ie, false negatives), particularly for drugs used as needed (eg, antacids, migraine medications).

The main goal in ascertaining exposure in any study is therefore to minimize the misclassification of the exposure status. Several approaches to operationalizing medication exposure during a time window in pregnancy based on filled prescriptions are possible (Figure 2), but each comes with tradeoffs. These include (1) days' supply of the medication overlapping the relevant exposure window, (2) at least one dispensed prescription during the relevant window, or (3) more than one dispensed prescription during the etiologically relevant window. Using days' supply, overlapping the exposure window will maximize the capture of relevant exposures (and thus study power) but has the potential to lead to misclassification if women stop the medication before the supply is exhausted. Using a dispensed prescription during the relevant exposure window likely increases the specificity for exposure, because most women probably take the medication immediately after it is dispensed, but this definition will lead to some decrement in power. Finally, exposure can be defined by requiring more than one prescription during the relevant exposure window. This reduces concerns regarding adherence or self-discontinuation, as it is likely that if a woman regularly refills a medication, then it is being taken as prescribed. This approach is only applicable to chronic treatments, which are also less affected by misclassification of the exposure window because those who fill multiple prescriptions will be using the medication over a longer period. Disadvantages to this approach include substantial loss of power for many exposures and potentially the selection of an unusual exposed population (particularly if the medication being studied is considered contraindicated in pregnancy or is not the preferred treatment for a given condition during pregnancy).

For many analyses, the preferred approach may be to define exposure based on a filled prescription during the relevant exposure window (approach 2) and then conduct sensitivity analyses using the other two approaches. Of note, it is also important to minimize misclassification of the nonexposed group. For example, if the exposure definition used is a dispensed prescription during the first trimester, women who were not dispensed a medication during the first trimester but who were dispensed the medication in the months prior to pregnancy should be excluded from the analysis. This is done to ensure that women who have a medication supply dispensed that might allow for use during the first trimester are not considered as unexposed. While this issue may have little impact when the unexposed general population is used as a reference because the probability of treatment for infrequently used medications among those without prescriptions would be minimal, when the reference group is restricted to women with an indication for treatment or when propensity score matching is used, the risk of including exposed women in the unexposed group (ie, false negatives) increases (since the reference group has a high probability of treatment).

5.2 | Etiologically relevant window

When defining the relationship between a medication and an adverse pregnancy outcome, it is critical to assess for exposure during the etiologically relevant exposure window. For example, if the outcome of interest is major congenital malformations, then the etiologically relevant exposure window is the first trimester when organogenesis occurs (for specific defects, this may be narrower; as an example, the relevant exposure window for the primary neurulation defects craniorachischisis, anencephaly, and open spina bifida ends 28 days after conception or 42 days after the LMP date³⁰). Or, in another example, if the outcome of interest is neonatal drug withdrawal, the etiologically relevant window would be used proximate to the time of delivery. Assessing for medication exposure during too broad a window or during the wrong window leads to exposure misclassification that will bias the relative risks pertaining to the sensitive window towards the null. It is rarely appropriate to define exposure based on medication use at any time point during pregnancy, as medication use is highly dynamic during pregnancy with medications frequently changed or discontinued by clinicians or patients because of real or perceived risks. Thus, it cannot be assumed that use at one time point during pregnancy implies use at other times.

While for some exposure-outcome associations (eg, congenital malformations) the etiologically relevant window is well defined, for others, where the underlying pathophysiology of the adverse outcome of interest is not fully understood or the potential biological relationship between the exposure and outcome not characterized, this may not be the case. For example, preeclampsia is thought to be caused by pathophysiological processes that occur at two stages in pregnancy. The first stage is because of abnormal placentation that occurs early in pregnancy; the second stage is because of endothelial function that occurs late in pregnancy. It is plausible that a medication can increase the risk of preeclampsia by impacting on either of these two mechanisms. Therefore, to assess the association between a medication and the risk of preeclampsia, both early and late pregnancy medication exposure should be evaluated.³¹ There may also be circumstances where the mechanism by which a drug confers excess risk is completely unknown; in this circumstance, the risks associated with exposure at different time points during pregnancy can be explored. Methods are currently under development to empirically define the relevant exposure window using scanning-based methods.³²

Evaluating the safety of medication with respect to different exposure windows can have important clinical implications that inform the way in which a medication is used during pregnancy or even in women of reproductive age who may become pregnant. For example, angiotensin-converting enzyme (ACE) inhibitor use during the second or third trimester of pregnancy is associated with a fetopathy characterized by renal atrophy, oligohydramnios, and pulmonary hypoplasia such that late pregnancy exposure to ACE inhibitors is clearly contraindicated.³³ Recent data suggest that early pregnancy exposure does not increase the risk for major congenital malformations after accounting for confounding factors.^{34,35} Thus, understanding the etiologically relevant window for exposure that causes fetal harm suggests that ACE inhibitors can likely be safely used in women in reproductive age, so long as they are stopped early in pregnancy.

5.3 | Accounting for preterm delivery

Not all pregnancies are 40 weeks in duration, and many outcomes of interest are associated with shorter gestational length (eg, preterm labor, preeclampsia, stillbirth). In those instances, it is important to avoid defining the exposure window in a way that creates differential opportunity for exposure in affected and unaffected pregnancies (eg, exposure “during the third trimester”) as this will bias the relative risk estimate. Several different strategies are available to avoid this bias, including (1) ending the exposure window prior to the start of follow-up (eg, at 24 weeks of gestation which is the earliest viable live birth) if exposure at that time is relevant, (2) defining exposure in a fixed lookback window from time of delivery (eg, exposure during the 30 days before delivery) assuming this exposure window is relevant and no trends of exposure towards the end of pregnancy, and (3) using a time-varying exposure definition (eg, exposure during week 36 is compared between preeclampsia cases and noncases within the risk set of women still pregnant at that time). The same principle would apply to exposure definitions during the first trimester when evaluating spontaneous abortions, ie, exposure to drugs during the first trimester would be less likely among pregnancies that end before the first trimester is complete.

5.4 | Dose-response analyses

An advantage of prescription claims databases is that they generally contain information on the strength of the medications (eg, in mg), the quantity of pills, and the days' supply dispensed allowing for the conduct of dose-response analyses. A clear dose response, which suggests a biological gradient of risk between the exposure and outcome, increases confidence that an observed association is likely to be causal (this is one of the Bradford Hill criteria³⁶). However, in situations where confounders are not well measured, a dose response may reflect an amplification of confounding by indication, as the confounding condition may be more severe among patients on higher doses. Defining the risks associated with different doses also has important clinical implications as it leads to a clearer understanding of the risks associated with a medication for a patient based on the dose she is taking, potentially altering the risk-to-benefit trade-off of use during pregnancy. For example, we examined the dose-response relation between first-trimester exposure to topiramate and the risks of oral clefts finding that the relative risk associated with doses less than or equal to 100 mg was 1.64 (95% CI 0.53-5.07), but for doses greater than 100 mg, it was 5.16 (95% CI 1.94-13.73).⁷ These findings suggest a biological dose response, but could be at least partially explained by the different indications for low and high doses (eg, migraine versus epilepsy). In addition to the daily dose, duration of treatment and/or cumulative dose during the etiologically relevant window may provide additional insights into the likelihood of a causal association.

5.5 | Selecting a reference group

Comparisons to unexposed women address questions about the safety of a drug versus no pharmacological treatment and no underlying indication (unexposed reference group). Comparisons to unexposed women with the same indication address questions about the safety of a drug versus no pharmacological treatment in women with the same indication, but potentially different severity (untreated condition reference group). Comparisons to

women exposed to other drugs for the same indication address questions about their comparative safety relative to alternative treatments (active treatment reference group), a clinically useful question. However, it is often challenging to identify a relevant active comparator that is widely used and for which definitive safety data are available. Comparisons to women who have been treated with the drug but discontinue early in pregnancy include women with the same indication and address questions about the safety of continuing treatment (discontinuers reference group).

5.6 | Bias analyses to account for exposure misclassification

It is important to quantify the potential impact of exposure misclassification because of unused dispensings, use of old prescriptions, or drug sharing. Aside from exploring alternative exposure definitions (Section 5.1), quantitative bias analyses can be conducted.^{37,38} Such analyses correct the relative risk for misclassification of the exposure based on user-specified probability distributions for the sensitivity and specificity of the exposure classification, referred to as the bias parameters. The bias parameters are repeatedly drawn from these distributions over many iterations (eg, 1000), and results are accumulated to generate a frequency distribution.¹⁰ A corrected relative risk estimate and simulation interval (defined by the 5th and the 95th percentile of the frequency distribution) can then be compared with the conventional relative risk estimate. While the impact of both nondifferential and differential exposure misclassification can be evaluated, it should be noted that the misclassification is likely to be nondifferential when using medication dispensings, which are recorded before the occurrence of the outcome.

6 | OUTCOME

6.1 | Selecting the outcomes of interest

Historically, drug safety evaluations in pregnancy have focused on birth defects. However, medications have effects that can impact on other birth outcomes including immunologic, vascular, hematologic, and metabolic effects. Certain drugs are of concern in pregnancy because they might amplify safety issues that exist outside of pregnancy (eg, because of changes in immune response or renal clearance), increase the risk of pregnancy-related complications (eg, preeclampsia), cause infant withdrawal after birth (eg, neonatal opioid withdrawal syndrome), or impact on neurocognitive development. Because healthcare utilization databases contain claims for procedures and other health services and the diagnoses used to justify these services, they can be used to study a wide range of maternal and infant outcomes, limited only by the range of diagnostic codes that are reliably recorded.³⁹

Investigators often face choices in defining the outcomes of interest. Selection of the study outcomes needs to be informed by the plausibility of their association with medication exposure, as well as their clinical significance. For example, analyses assessing the association between medication exposure and congenital malformations often focus on major structural defects. These are generally defined as abnormalities in structural development that are medically or cosmetically significant, are present at birth, and persist in postnatal life unless or until repaired.^{40,41} This definition excludes minor anomalies such as

small muscular ventricular septal defects that may spontaneously close with no consequences for the infant and defects that are transient and maturational such as a patent ductus arteriosus that might occur in a preterm infant. Malformations of known etiology (eg, chromosomal, because of known teratogens) are often excluded since these are not attributable to the exposure and their inclusion can decrease the likelihood of detecting an exposure-related risk.

Attaining sufficient study size when studying the effects of medications in pregnancy is difficult because exposures are often uncommon, and outcomes are often rare. As a result, investigators often revert to grouping outcomes, in particular malformations. For example, a common primary outcome used in safety studies is overall congenital malformations, which are diagnosed in 2% to 3% of live born infants; while a specific malformation (eg, oral clefts) may have a prevalence of about 1 per 1000 live births. This is, however, a potentially problematic solution, as assessing the risk of a compound outcome like malformations overall may mask an association with rare, but clinically consequential, malformations.⁴² Thus, when possible, analyses should assess for the risk of individual malformations, organ-specific malformations, or malformations that share common embryological origins. However, absent power to examine these, defining the risk for malformations overall is still clinically meaningful in order to identify whether a medication is a major teratogen (eg, thalidomide's preferential effect on limb defects would result in an elevated risk of malformations overall).

Investigators also need to determine when outcomes are going to be ascertained. Some pregnancy events are acute, and their window of ascertainment is straightforward (eg, premature rupture of membranes, cesarean section), while others are progressive and have less well-defined windows of ascertainment (eg, neurocognitive development).

6.2 | Using databases to ascertain outcomes

Ascertaining pregnancy outcomes using codes in claims is not trivial, and algorithms need to be carefully developed based on diagnoses and procedures and then validated. Although it is common practice, for many outcomes, it is not appropriate to use a single diagnostic code recorded in claims as evidence of the presence of a condition because single codes are often recorded to justify the performance of a diagnostic test, which may rule out the suspected diagnosis. Algorithms that identify the outcome of interest with high specificity, even at the cost of some sensitivity, should be developed and validated based on chart review because if an outcome is defined with near perfect specificity, and nondifferential sensitivity, then relative risk estimates will be unbiased.⁴³

For example, in MAX, if major congenital malformations are defined on the basis of a single diagnostic code, the prevalence of malformations exceeds 10%, about three times higher than is expected in the US population based on other sources.⁴⁴ Defining malformations in this manner leads to a high degree of misclassification, which would bias risk estimates towards the null, assuming that the misclassification is nondifferential. In contrast, defining major malformations based on either (1) more than one date with a code indicating a particular malformation, (2) at least one date plus a corrective surgery or procedure for that malformation, or (3) at least one date plus infant death within 30 days results in a prevalence

of major congenital malformations that is in line with expectation (Figure 3). Further, the validity of such approach has been demonstrated based on validation studies using medical charts^{45,46} and, indirectly, by replicating the association between a known risk factor (ie, prepregnancy diabetes) and malformations defined using this algorithm.¹⁴ For several outcomes, definitions with a high Positive Predictive Value (PPV) have been developed (Table 1).

6.3 | Bias analyses to account for outcome misclassification

Even when using highly specific outcome definitions in claims, some misclassification is expected. Most concerning is the possibility of differential misclassification of the outcome with respect to the exposure in certain settings. For example, diagnostic bias may occur if the exposed group receives more prenatal or perinatal testing (eg, because of suspicion of teratogenic effects) that results in more complete diagnosis, or over diagnosis of subclinical anomalies. The bias introduced by such misclassification can be modeled using probabilistic bias analysis^{37,38} based on either a PPV estimated from an internal validation study or using a range of plausible PPVs. In this approach, a minimum, mode, and maximum for sensitivity and specificity are specified, consistent with the PPV (see Section 5.6). In practice, with a high PPV and negative predictive value (NPV) close to 100%, the percentage change from the conventional estimate is expected to be small.

6.4 | Pregnancy losses, identification, and characterization

The occurrence of embryonic or fetal demise raises two distinct concerns. First, healthcare utilization databases can under-record spontaneous or induced pregnancy losses that do not end up with a health encounter within the system. If the missingness is random with respect to exposure, this limitation will affect the absolute risk of pregnancy losses, but the relative risk will be unbiased. Nothing can be done other than quantifying the underestimation based on published data from birth and fetal death certificates. Second, when the goal is to study malformations, right truncation occurs even when losses are captured since follow-up ends with unknown outcomes because not all fetuses undergo autopsy and, when they do, it is unlikely for the pathology results to be coded in claims as it is not billable information. For this reason, most cohorts consider the prevalence of malformations in live births rather than the cumulative incidence over gestation. Failure to include defects detected among terminations can underestimate the incidence of malformations and decrease precision of the estimate, particularly for malformations frequently resulting in termination after prenatal diagnosis (eg, neural tube defects), but if the missingness is random with respect to exposure, the relative risk for malformations will be unbiased (see Section 8).

7 | CONFOUNDING

7.1 | Confounding by indication and other risk factors

Women treated with a medication during pregnancy often differ in important ways from women who are not. Foremost, they differ with respect to the presence and severity of the underlying condition for which the treatment is being prescribed. If this treatment indication or factors associated with it are also risk factors for the study outcome, then confounding is present. Such confounding by indication is widely recognized as a major threat to the

validity of pharmacoepidemiologic studies but has traditionally received scant attention in pregnancy research, likely because the focus tends to be on unintended treatment effects where the potential for confounding by indication is less. However, although many treatment indications (eg, psychiatric illness) might not directly increase the risk for certain adverse pregnancy outcomes, they frequently serve as proxies for associated conditions and behaviors that do increase the risk of the outcome. For example, smoking, alcohol and illicit drug use, poor maternal diet, obesity, suboptimal prenatal care, and chronic conditions such as diabetes and hypertension are all more common in patients with psychiatric illness and are potential risk factors for congenital cardiac anomalies.⁴⁷ If the psychiatric illness is not accounted for, the observed association with the outcome will be a mixture of the effect of the medication itself and the effect of the condition and its associated factors.

Aside from the treatment indication, there are numerous other potential risk factors for the maternal or neonatal outcome under consideration that may be imbalanced between exposure groups. While clinical characteristics are well documented in healthcare utilization data, these administrative data sources typically do not contain information on inpatient medication use or medications available over the counter (eg, multivitamins and folate), and they lack robust information on behavioral and lifestyle factors such as prepregnancy weight, body mass index, smoking, illicit drug use, alcohol use, socioeconomic status, and familial history.

As is the case for exposure and outcome, the ascertainment window for covariates needs to be specified. A general principle when used for confounding control is to measure covariates before the outcome occurs (to avoid selection bias and recording bias).^{48,49} For example, for studies of teratogenesis, covariates are generally measured during a prepregnancy baseline period through the end of the first trimester to avoid measuring covariates that might occur as a consequence of malformations or preferential recording after a malformation is recognized by prenatal screening. While covariates are also typically measured before the exposure window (to avoid adjusting for factors in the causal pathway), this temporality can be waived under certain assumptions. For example, a chronic condition likely to predate conception can be ascertained during pregnancy to increase the opportunity to capture the diagnoses as health encounters accumulate over time. Also, conditions that arise late in pregnancy may be good proxies for risk factors that are not well recorded (eg, a diagnosis of gestational diabetes which typically occurs after 24 gestational weeks may be a good proxy for obesity at the start of pregnancy which is known to be under-recorded in healthcare utilization data).

7.2 | Design and analytic approaches to address confounding

Because of the lack of randomization in nonexperimental study designs, there is no intrinsic mechanism that ensures balance in patient characteristics between the treated and the reference group. Several design and analytic approaches that take advantage of the richness of the administrative data can be implemented, however, to help reduce potential confounding.

7.2.1 | Design approaches to address confounding by indication—First, both the treated and the untreated population can be restricted to women with a recorded diagnosis for the underlying indication. Although this will substantially reduce the study size, it will increase the internal validity, and the effect on precision tends to be relatively modest since such restriction affects mostly the larger untreated group. For example, in a study evaluating the risk of congenital cardiac malformations associated with first-trimester exposure to selective serotonin reuptake inhibitors (SSRIs), restricting the population to women with a recorded diagnosis of depression without any further adjustment markedly attenuated the associations.¹⁰ Alternatively, rather than restricting the population, one can adjust for the underlying condition by including it in the propensity score (PS), for example, and trimming patients in the nonoverlapping regions of the PS distribution to reduce the likelihood of nonpositivity regarding the indication.^{7,50} The later approach trusts the application of important inclusion criteria to the trimming, resulting in less well-defined criteria more difficult to replicate, but in our experience, restriction and trimming usually lead to similar results in practice. Please refer to Section 7.2.2 for further details on confounding adjustment using PS.

Another possible approach is to compare the risk of the outcome between women who were treated with the medication of interest before the start of pregnancy and continued versus discontinued during the etiologically relevant exposure time window.²³ Finally, one can select women treated with an alternative medication (ie, active comparator) as the reference group rather than untreated women to maximize the equipoise (see Section 5.5).

7.2.2 | Analytic approaches to account for measured confounders—The richness of the information included in healthcare utilization databases can be exploited to identify a large number of potential risk factors for the outcome or proxies for them. Examples of proxies include general markers of the burden of illness, such as the obstetric comorbidity index,⁵¹ the number of distinct prescriptions for medications other than the exposure (and comparator) of interest, number of distinct diagnoses, number of outpatient visits, hospitalizations, and emergency room visits.

To assess the exchangeability of exposed and reference groups, the first step in the analysis is to compare the distributions of sociodemographic, clinical, and healthcare utilization characteristics during the relevant baseline period for the exposed and reference group. Balance can be assessed using the standardized mean difference. Typically, an absolute standardized difference greater than 0.1 is considered an indicator for imbalances between the two groups.⁵²

Given that adverse pregnancy outcomes tend to be rare and the set of prespecified potential confounding variables may be large, the use of data reduction techniques such as PS will often be advisable to avoid problems with model overfitting.⁵³ Briefly, PS represents the predicted probability of treatment conditional on observed characteristics and, in their simplest form, are estimated in a logistic regression model. Confounding can then be controlled through matching, stratification, weighting, or covariate adjustment using the PS.⁵⁴ It has been demonstrated that especially fine PS stratification and weighting approaches offer some advantages over matching in the context of pregnancy research, where exposure

to drugs is often infrequent and the PS distribution tends to be right skewed (ie, a large proportion of the cohort has low probability of receiving the treatment). Fine stratification using the PS distribution among the exposed provides equivalent confounding control as matching with greater precision.⁵⁵ In such approach, n equally sized PS strata are created based on the distribution among the treated women, after trimming the cohort by excluding observations from the nonoverlapping regions of the PS distributions. The outcome models can either be stratified, or the untreated observations can be weighted using the distribution of the treated among the PS strata. The effect estimate will reflect the average treatment effect on the treated (ATT estimate). The advantage of the weighting approach is that it can also be used to present the balance in the baseline characteristics after accounting for potential confounding variables in descriptive Table 1 (similar to 1:1 matching).⁵⁵ It should be noted, however, that this approach assumes that the treated group is much smaller than the reference group. If this is not the case, then stratification based on the distribution of the entire study population (rather than the treated women) among the PS strata is preferred to avoid sparse data problems, in which case the effect estimate represents the average treatment effect (ATE estimate). Note that the ATE estimate would be interpretable only if the study population had been restricted to those eligible for treatment in clinical practice, both the treated and the reference groups. Finally, since PS can only adjust for baseline covariates, for time-varying confounding methods such as inverse probability, weighting can be used.⁵⁶

7.2.3 | Analytic approaches to account for unmeasured confounding—As in any pharmacoepidemiologic study, there is always concern about residual confounding by unmeasured or poorly measured variables. We can make use of a broad array of approaches to address concerns about residual confounding: (1) internal adjustment approaches, (2) external adjustment approaches, and (3) sensitivity analyses that test the robustness of the findings to residual confounding.

1. Internal adjustment approaches: Given that most data sources lack robust information on one or more potential confounders, it might be prudent to always conduct a confirmatory analysis using high-dimensional PS.⁵⁷ The high-dimensional PS algorithm evaluates thousands of diagnoses, procedures, and pharmacy claim codes in order to identify and prioritize those covariates that may serve as proxies for unmeasured confounders. The generous inclusion of these empirically identified confounders—often several hundreds of them—combined with the investigator-identified prespecified covariates has been shown to improve confounding control in some circumstances. Figure 4 illustrates the effect of increasing levels of confounding adjustment on the association between concomitant exposure to antipsychotics and prescription opioids versus prescription opioids alone and the risk of neonatal abstinence syndrome.

Another useful approach is to conduct a sibling discordance study. Outcomes are compared among siblings who are discordant for the exposure of interest. This design exploits the fact that siblings share a family environment and genetic factors. Any remaining differences between siblings are potential confounders which can be controlled in the analyses if measurable. The crucial assumption underlying this design is that family factors relevant to

the outcome are stable (eg, genetic predisposition to both depression and autism) and therefore similar across siblings. Sibling discordance studies are particularly valuable as a sensitivity analysis to confirm a potentially causal association identified in a study of unrelated individuals.^{59,60} The main challenge for this design is the need for a sufficient number of discordant siblings. The reduced sample size often results in unstable estimates; confounding by acute indications is not controlled and might be stronger, and the likelihood of systematic exposure misclassification is higher.⁶¹

Instrumental variable (IV) analysis is a technique that allows adjustment for unmeasured as well as measured risk factors. An instrument can be defined as a variable that is independent of the unmeasured confounding, affects treatment decisions, and has no effect on the outcome independent from its effect on the treatment.⁶² Studying the association between the instrument and the outcome while adjusting for the association between the instrument and the treatment will result in an unbiased effect estimate if the conditions for a valid IV are met. The challenges of this approach are that some of the conditions for a valid IV are untestable, and weak instruments result in imprecise and sometimes inconsistent estimates.⁶³ While there are some examples of IV analyses in perinatal epidemiology⁶⁴⁻⁷⁰—several of them using Mendelian randomization—the approach has not been extensively explored in perinatal pharmacoepidemiology.^{71,72} In an attempt to validly estimate the effectiveness of antidepressants during pregnancy, Swanson et al evaluated preference-based, calendar-time-based, and geography-based instruments, but concluded that they did not meet the required conditions for a valid analysis: Several of the tested instruments were weak, and all failed to provide better balance of measured covariates than the actual exposure status.⁷¹

It is worth noting that the identification of potential confounders needs to be informed by subject matter expertise.⁷³ Including variables into the model just because they are in the data can make bias worse by introducing selection bias. A classic example is the inclusion of birth weight or preterm birth as covariates.⁴⁸

2. External adjustment approaches: If information on the potential confounding variables is not available in the main dataset, but is available in other datasets or can be collected for a subset of the study patients (through linkage with electronic health records or through a survey), external adjustment of unmeasured confounders using techniques such as PS calibration or two-stage designs becomes a useful strategy to gain insight into the magnitude of potential residual bias.^{74,75} One useful data source for such external adjustment in the United States is the National Health and Nutrition Examination Survey (NHANES), which is an ongoing nationally representative health examination survey conducted by the National Center for Health Statistics of the Centers for Disease Control and Prevention^{76,77} to assess the health of the population. NHANES collects information on lifestyle factors, such as smoking and obesity, which are known to be under-recorded in healthcare utilization data. The NHANES has been previously used to describe the prevalence of diseases and risk factors, set national standards for health indicators, and evaluate their time trends.

3. Sensitivity analyses to test robustness: Using algebraic methods,⁷⁸ one can study the impact of an unmeasured confounder and estimate an externally adjusted relative risk

(RR_{ED}) relating exposure (E) and disease (D) within levels of the confounder (C). Graphs that present a range of plausible values for the externally adjusted (ie, stratified) RR_{ED} as a function of the prevalence of the confounder in the reference group (P_0) and in the exposed (P_1), the association between the confounder and the disease (RR_{CD}), and the unadjusted RR_{ED} can be created (Figure 5). The stratified (ie, externally adjusted) RR_{ED} can be estimated as $RR_{ED} \times P_0 / P_1 \times [RR_{CD} + (1 / P_0) - 1] / [RR_{CD} + (1 / P_1) - 1]$. Rather than a graphical (or tabular) display of the externally adjusted RR_{ED} for a range of plausible values of the bias parameters P_0 , P_1 , and RR_{CD}, probabilistic sensitivity analyses provide a point estimate with a 95% simulation interval, which combines both the random and systematic error.^{37,38}

Alternatively, one can assess the strength of the unmeasured confounding that would be required to fully explain the treatment-outcome association. This can be done using a graphical or tabular approach that illustrates the various combinations of the strength of the association between the unmeasured confounder and the exposure and outcome, respectively,⁷⁸⁻⁸⁰ or by using the *E* value approach.⁸¹ In addition to showing the minimum strength of association that an unmeasured confounder would need to have to explain a positive association, (non-null) *E* values can also be helpful to evaluate the minimum magnitude of both confounder associations needed to shift an estimate near the null to one that is clinically meaningful (similar to the externally adjusted risk).⁸¹ We previously reported an adjusted risk ratio estimate of 1.06 (95% CI, 0.93-1.22) for the use of SSRI antidepressants during pregnancy on infant cardiac defects.¹⁰ A hypothetical unmeasured confounder that was associated with both SSRI antidepressant use and infant cardiac defects by a risk ratio of 1.5-fold each could move the risk ratio to 1.20, but with weaker confounding could not.⁸¹

Negative control analyses can be helpful to diagnose the presence of residual confounding. For example, if a positive association is found between drug exposure during the first trimester and the risk of congenital malformations but residual confounding is a concern, the risk of malformations can be assessed in women who filled their first prescription for the medication of interest outside of the etiologically relevant window. A null finding in this analysis provides indirect evidence of no substantial residual confounding; a positive association suggests that residual confounding is highly likely.¹⁵ Alternatively, one can assess the association between first-trimester exposure and a different outcome that is not expected to be associated with first-trimester exposure, but is likely to be equally confounded if residual confounding is present (negative control outcome). A null finding would again provide reassurance that the results are unlikely to be explained by residual confounding whereas a positive association would strengthen the concern. Paternal exposures can also be used as a negative control. Any association between paternal treatments during their infant's gestation period with pregnancy outcomes would be expected to be because of familial confounding, assuming that paternal exposure does not have a direct biological effect.

8 | SELECTION BIAS

Claims-based studies of drug safety are often restricted to live births because nonlive birth pregnancy outcomes (ie, therapeutic and spontaneous abortions, stillbirths) may not be well captured in claims databases, coupled with the challenges in estimating the LMP for these outcomes and the infrequent recording of fetal malformations if identified. This can result in right truncation, in which follow-up ends before an outcome is detected (eg, miscarriage before a congenital malformation can be identified at birth). As such, this has the potential to introduce selection that may bias the drug exposure-outcome association being assessed. Methods have therefore been developed to quantify the potential bias associated with restriction to live births.^{10,14,82,83}

If the proportion of pregnancies ending with a live birth is the same in the exposed and reference groups within the levels of the covariates adjusted for in the analyses, then the relative risk estimates derived from the live birth–restricted cohort will be less likely to be biased. If, however, nonlive birth outcomes occur more frequently in the exposed group, then relative risk estimates for congenital malformations from the live birth–restricted analysis might be biased towards the null. Certain medication exposures might be associated with more frequent prenatal care and screening and/or a higher propensity to terminate pregnancy if a malformation is diagnosed. The impact of differential live birth frequencies by exposure and outcome can be modeled using a range of assumptions informed by the literature and corrected relative risk estimates generated (see Figure 6 for details).

It is important not to confuse selection bias with competing risks. In the presence of competing risks (eg, abortions with stillbirths, preterm delivery with preeclampsia), having one outcome prevents another outcome later in pregnancy. In the case of selection bias, the outcome had occurred, but we did not observe it (eg, when evaluating teratogenicity among live births). When we have competing risks, the second outcome will not occur; in a sense, it is a real reduction of the risk for one outcome mediated through having another outcome.^{84,85}

9 | EXTERNAL VALIDITY

Even though administrative databases typically include a large unselected population, inevitably questions arise regarding the generalizability of the study findings. For example, can findings from a study conducted in publicly insured women be generalized to privately insured women who might differ in terms of socioeconomic status, age, racial composition, and morbidity burden? It is important to remember that scientific generalization of valid estimates of effect (ie, external validity) does not require representativeness of the study population in a survey-sampling sense (like a random sample of patients is not required in clinical trials). The relevant question is whether the distinguishing characteristics of the studied groups modify the effect of interest.⁴³ In the context of pregnancy cohorts, generalizability will therefore depend on whether the biologic relations studied are expected to differ between women included in the study and those that are not. Knowledge from different scientific disciplines—not just epidemiology—will contribute to this assessment, and to some extent, this can be verified in the data. Analyses can be stratified by those

factors that are believed to distinguish the study population from the target populations and are suspected to be effect modifiers. If there is no evidence of effect measure modification by these factors, results are likely to be generalizable.¹⁰ If there is, then it is important to present the stratified results to allow standardization of the effect to other populations. For example, thalidomide was a teratogen worldwide; there was no need to confirm effects in all continents. However, the preventive effects of folate supplements might depend on the dietary intake of folic acid in different populations.

The distribution of characteristics, as well as weights to standardize estimates to the distributions in the overall US obstetric population, can be obtained from representative samples such as the nationwide inpatient sample (NIS), a 20% stratified sample of all US community hospitals, and NHANES. For example, we confirmed with the NIS that 45% of the delivery hospitalizations are covered by Medicaid,⁸⁶ the distribution of maternal characteristics in the NIS Medicaid population was similar to that observed in MAX, and crude associations between risk factors and outcomes were consistent.⁸⁷

10 | MISUSE OF STATISTICAL SIGNIFICANCE

Because the focus of epidemiologic research is on estimation and not on significance testing, adjustment for multiple testing is not recommended.⁸⁸⁻⁹⁰ Nevertheless, as for any study, it is imperative to prespecify the analyses in the study protocol and to report all findings regardless of the magnitude of the effects and their statistical significance. *P* value functions are a useful way to help with the interpretation of the strength of the evidence provided by the study, as they visualize two key components of the estimation process: the magnitude of the effect and the precision with which it is estimated.⁸⁸ Study findings should always be interpreted in the context of the totality of the evidence that has accumulated over time. This can be accomplished indirectly through careful discussion of the evidence from previous reports or more directly through the conduct of a formal meta-analysis as part of the new study¹¹ or through the use of Bayesian approaches.⁹¹ Any new unexpected finding should be viewed as an initial safety signal that requires confirmation in additional studies,⁸ and researchers should avoid the temptation to make it the focus of the report.

11 | CONCLUSION

Healthcare utilization data provide a powerful data source for studying drug safety in pregnancy. The value of these data is increasingly recognized by regulatory authorities as a useful complement to the approaches that have been the foundation for pharmacoepidemiological studies in pregnancy: registries and case-control designs (Table 2). However, there are a number of important methodological considerations that need to be attended to in order to generate valid, minimally biased risk estimates. The impact of these data sources in the field of drug safety in pregnancy promises to be further enhanced as partnerships—such as the Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP)^{39,92} or the International Pregnancy Safety Study (InPreSS) consortium^{8,12}—are formed that pool healthcare utilization data drawn from different populations to define with greater precision and generalizability the risks associated with particular drugs. Work based on healthcare utilization databases is likely to be strengthened in the future as linkages are

established to external databases with additional clinical information (eg, birth certificates, laboratory tests, electronic medical records) that can be used for confounding control and outcome ascertainment.

Acknowledgments

FUNDING

This study is financially supported by the National Institute of Child Health and Human Development (R21HD092879-01), the National Institute of Mental Health (R01MH116194-01), and the National Institute on Drug Abuse (R01DA044293-01A1).

Funding information

National Institute of Child Health and Human Development, Grant/Award Number: R21HD092879-01; National Institute on Drug Abuse, Grant/Award Number: R01DA044293-01A1; National Institute of Mental Health, Grant/Award Number: R01MH116194-01

REFERENCES

- Heyrana K, Byers HM, Stratton P. Increasing the participation of pregnant women in clinical trials. *JAMA*. 11 27 2018;320(20):2077–2078. [PubMed: 30422300]
- Hernandez-Diaz S. Pregnancy registries. In: AHRQ, ed. *Registries for Evaluating Patient Outcomes: A User's Guide: 3rd Edition* 2014:135–168.
- Food and Drug Administration. Guidance for industry: establishing pregnancy exposure registries. 8 2002 (<https://http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm071639.pdf>).
- Mitchell AA. Systematic identification of drugs that cause birth defects —a new opportunity. *N Engl J Med*. 2003;349(26):2556–2559. [PubMed: 14695418]
- Palmsten K, Huybrechts KF, Mogun H, et al. Harnessing the Medicaid Analytic eXtract (MAX) to evaluate medications in pregnancy: design considerations. *PLoS ONE*. 2013;8(6):e67405. [PubMed: 23840692]
- Markus AR, Andres E, West KD, Garro N, Pellegrini C. Medicaid covered births, 2008 through 2010, in the context of the implementation of health reform. *Women's Health Issues: Official Publication of the Jacobs Institute of Women's Health*. Sep-Oct 2013;23(5):e273–e280.
- Hernandez-Diaz S, Huybrechts KF, Desai RJ, et al. Topiramate use early in pregnancy and the risk of oral clefts: a pregnancy cohort study. *Neurology*. 1 23 2018;90(4):e342–e351. [PubMed: 29282333]
- Huybrechts KF, Broms G, Christensen LB, et al. Association between methylphenidate and amphetamine use in pregnancy and risk of congenital malformations: a cohort study from the International Pregnancy Safety Study Consortium. *JAMA Psychiat*. 2 1 2018;75(2):167–175.
- Paterno E, Huybrechts KF, Bateman BT, et al. Lithium use in pregnancy and the risk of cardiac malformations. *N Engl J Med*. 6 8 2017;376(23):2245–2254. [PubMed: 28591541]
- Huybrechts KF, Palmsten K, Avorn J, et al. Antidepressant use in pregnancy and the risk of cardiac defects. *N Engl J Med*. 6 19 2014;370(25):2397–2407. [PubMed: 24941178]
- Huybrechts KF, Bateman BT, Palmsten K, et al. Antidepressant use late in pregnancy and risk of persistent pulmonary hypertension of the newborn. *JAMA*. 6 2 2015;313(21):2142–2151. [PubMed: 26034955]
- Bateman BT, Heide-Jorgensen U, Einarsdottir K, et al. Beta-blocker use in pregnancy and the risk for congenital malformations: an international cohort study. *Ann Intern Med*. 11 20 2018;169(10):665–673. [PubMed: 30326014]
- Bateman BT, Paterno E, Desai RJ, et al. Late pregnancy beta blocker exposure and risks of neonatal hypoglycemia and bradycardia. *Pediatrics*. 9 2016;138(3):e20160731. [PubMed: 27577580]

14. Bateman BT, Hernandez-Diaz S, Fischer MA, et al. Statins and congenital malformations: cohort study. *BMJ*. 3 17 2015;350(mar17 10): h1035. [PubMed: 25784688]
15. Huybrechts KF, Hernandez-Diaz S, Straub L, et al. Association of maternal first-trimester ondansetron use with cardiac malformations and oral clefts in offspring. *JAMA*. 12 18 2018;320(23):2429–2437. [PubMed: 30561479]
16. Desai RJ, Huybrechts KF, Hernandez-Diaz S, et al. Exposure to prescription opioid analgesics in utero and risk of neonatal abstinence syndrome: population based cohort study. *BMJ*. 5 14 2015;350(may14 1):h2102. [PubMed: 25975601]
17. Hernandez-Diaz S, Hernan MA, Meyer K, Werler MM, Mitchell AA. Case-crossover and case-time-control designs in birth defects epidemiology. *Am J Epidemiol*. 8 15 2003;158(4):385–391. [PubMed: 12915504]
18. Andrade SE, Toh S, Houstoun M, et al. Surveillance of medication use during pregnancy in the mini-sentinel program. *Matern Child Health J*. 4 2016;20(4):895–903. [PubMed: 26645616]
19. Johnson KE, Beaton SJ, Andrade SE, et al. Methods of linking mothers and infants using health plan data for studies of pregnancy outcomes. *Pharmacoepidemiol Drug Saf*. 7 2013;22(7):776–782. [PubMed: 23596095]
20. Ananth CV, Platt RW, Savitz DA. Regression models for clustered binary responses: implications of ignoring the intracluster correlation in an analysis of perinatal mortality in twin gestations. *Ann Epidemiol*. 4 2005;15(4):293–301. [PubMed: 15780777]
21. Oberg AS, VanderWeele TJ, Almqvist C, Hernandez-Diaz S. Pregnancy complications following fertility treatment-disentangling the role of multiple gestation. *Int J Epidemiol*. 8 1 2018;47(4): 1333–1342. [PubMed: 29939263]
22. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158(9):915–920. [PubMed: 14585769]
23. Park Y, Hernandez-Diaz S, Bateman BT, et al. Continuation of atypical antipsychotic medication during early pregnancy and the risk of gestational diabetes. *Am J Psychiatry*. 6 1 2018;175(6):564–574. [PubMed: 29730938]
24. Margulis AV, Palmsten K, Andrade SE, et al. Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results. *Pharmacoepidemiol Drug Saf*. 4 2015;24(4):335–342. [PubMed: 25627986]
25. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf*. 1 2013;22(1):16–24. [PubMed: 22550030]
26. Li Q, Andrade SE, Cooper WO, et al. Validation of an algorithm to estimate gestational age in electronic health plan databases. *Pharmacoepidemiol Drug Saf*. 5 2013;22(5):524–532. [PubMed: 23335117]
27. Hornbrook MC, Whitlock EP, Berg CJ, et al. Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system. *Health Serv Res*. 4 2007;42(2):908–927. [PubMed: 17362224]
28. Honein MA, Paulozzi LJ, Cragan JD, Correa A. Evaluation of selected characteristics of pregnancy drug registries. *Teratology*. 1999;60(6):356–364. [PubMed: 10590397]
29. Cragan JD, Gilboa SM. Including prenatal diagnoses in birth defects monitoring: experience of the metropolitan Atlanta congenital defects program. *Birth Defects Res A Clin Mol Teratol*. 2009;85(1):20–29. [PubMed: 19089857]
30. Greene ND, Copp AJ. Neural tube defects. *Annu Rev Neurosci*. 2014;37(1):221–242. [PubMed: 25032496]
31. Cohen JM, Hernandez-Diaz S, Bateman BT, et al. Placental complications associated with psychostimulant use in pregnancy. *Obstet Gynecol*. 12 2017;130(6):1192–1201. [PubMed: 29112657]
32. Huybrechts KF, Mogun H, Bateman BT, et al. Detecting the etiologically relevant risk window for medications used in pregnancy. *Pharmacoepidemiol Drug Saf*. 8 2017;26:469–470. [PubMed: 28220982]
33. Nadeem S, Hashmat S, Defreitas MJ, et al. Renin angiotensin system blocker fetopathy: a midwest pediatric nephrology consortium report. *J Pediatr*. 10 2015;167(4):881–885. [PubMed: 26130112]

34. Bateman BT, Paterno E, Desai RJ, et al. Angiotensin-converting enzyme inhibitors and the risk of congenital malformations. *Obstet Gynecol.* 1 2017;129(1):174–184. [PubMed: 27926639]
35. Li DK, Yang C, Andrade S, Tavares V, Ferber JR. Maternal exposure to angiotensin converting enzyme inhibitors in the first trimester and risk of malformations in offspring: a retrospective cohort study. *BMJ.* 10 18 2011;343(oct18 1):d5931.
36. Ward AC. The role of causal criteria in causal inferences: Bradford Hill's “aspects of association”. *Epidemiologic perspectives & innovations: EP+I.* 6 17 2009;6:2. [PubMed: 19534788]
37. Lash T, Fox M, Fink A. *Applying quantitative bias analysis to epidemiologic data.* New York: Springer; 2009.
38. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol.* 12 2005;34(6):1370–1376. [PubMed: 16172102]
39. Andrade SE, Scott PE, Davis RL, et al. Validity of health plan and birth certificate data for pregnancy research. *Pharmacoepidemiol Drug Saf.* 1 2013;22(1):7–15. [PubMed: 22753079]
40. Holmes LB, Westgate MN. Inclusion and exclusion criteria for malformations in newborn infants exposed to potential teratogens. *Birth Defects Res A Clin Mol Teratol.* 9 2011;91(9):807–812. [PubMed: 21800414]
41. Holmes LB, Westgate MN. Using ICD-9 codes to establish prevalence of malformations in newborn infants. *Birth Defects Res A Clin Mol Teratol.* 4 2012;94(4):208–214. [PubMed: 22451461]
42. Khoury MJ, Moore CA, James LM, Cordero JF. The interaction between dysmorphology and epidemiology: methodologic issues of lumping and splitting. *Teratology.* 1992;45(2):133–138. [PubMed: 1615423]
43. Rothman K, Greenland S, Lash T. Validity in epidemiologic studies In: Rothman K, Greenland S, Lash T, eds. *Modern Epidemiology.* third ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
44. Centers for Disease C, Prevention. Update on overall prevalence of major birth defects—Atlanta, Georgia, 1978–2005. *MMWR Morb Mortal Wkly Rep.* 1 11 2008;57(1):1–5. [PubMed: 18185492]
45. Palmsten K, Huybrechts KF, Kowal MK, Mogun H, Hernandez-Diaz S. Validity of maternal and infant outcomes within nationwide Medicaid data. *Pharmacoepidemiol Drug Saf.* 6 2014;23(6): 646–655. [PubMed: 24740606]
46. He M, Cottrill JA, Bartels DD, et al. Validation of algorithms to identify perinatal outcomes in large claims databases. *Pharmacoepidemiol Drug Saf.* 8 2017;26:98–99.
47. Jenkins KJ, Correa A, Feinstein JA, et al. Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation.* 6 12 2007;115(23):2995–3014. [PubMed: 17519397]
48. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight “paradox” uncovered. *Am J Epidemiol.* 12 1 2006;164(11):1115–1120. [PubMed: 16931543]
49. MacDonald SC, Hernan MA, McElrath TF, Hernandez-Diaz S. Assessment of recording bias in pregnancy studies using health care databases: an application to neurologic conditions. *Paediatr Perinat Epidemiol.* 5 2018;32(3):281–286. [PubMed: 29569366]
50. Huybrechts KF, Hernandez-Diaz S, Paterno E, et al. Antipsychotic use in pregnancy and the risk for congenital malformations. *JAMA Psychiat.* 9 1 2016;73(9):938–946.
51. Bateman BT, Mhyre JM, Hernandez-Diaz S, et al. Development of a comorbidity index for use in obstetric patients. *Obstet Gynecol.* 11 2013;122(5):957–965. [PubMed: 24104771]
52. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput.* 2009;38(6):1228–1234.
53. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol.* 3 2006;98(3):253–259. [PubMed: 16611199]
54. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res.* 5 2011;46(3):399–424.

55. Desai RJ, Rothman KJ, Bateman BT, Hernandez-Diaz S, Huybrechts KF. A propensity-score-based fine stratification approach for confounding adjustment when exposure is infrequent. *Epidemiology* 3 2017;28(2):249–257. [PubMed: 27922533]
56. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 12 10 2015;34(28):3661–3679. [PubMed: 26238958]
57. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 7 2009;20(4):512–522. [PubMed: 19487948]
58. Huybrechts KF, Bateman BT, Desai RJ, et al. Risk of neonatal drug withdrawal after intrauterine co-exposure to opioids and psychotropic medications: cohort study. *BMJ* 2017;358:j3326. [PubMed: 28768628]
59. Donovan SJ, Susser E. Commentary: advent of sibling designs. *Int J Epidemiol* 4 2011;40(2):345–349. [PubMed: 21450688]
60. Oberg AS, D'Onofrio BM, Rickert ME, et al. Association of labor induction with offspring risk of autism spectrum disorders. *JAMA Pediatr*. 9 6 2016;170(9):e160965. [PubMed: 27454803]
61. Keyes KM, Smith GD, Susser E. On sibling designs. *Epidemiology*. 5 2013;24(3):473–474. [PubMed: 23549193]
62. Guo Z, Cheng J, Lorch SA, Small DS. Using an instrumental variable to test for unmeasured confounding. *Stat Med*. 9 10 2014;33(20):3528–3546. [PubMed: 24930696]
63. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 7 2006;17(4):360–372. [PubMed: 16755261]
64. Austin N, Harper S, Strumpf E. Does segregation lead to lower birth weight?: an instrumental variable approach. *Epidemiology*. 9 2016;27(5):682–689. [PubMed: 27182641]
65. Murray J, Burgess S, Zuccolo L, Hickman M, Gray R, Lewis SJ. Moderate alcohol drinking in pregnancy increases risk for children's persistent conduct problems: causal effects in a Mendelian randomisation study. *J Child Psychol Psychiatry*. 5 2016;57(5):575–584. [PubMed: 26588883]
66. Aubrey-Bassler K, Cullen RM, Simms A, et al. Outcomes of deliveries by family physicians or obstetricians: a population-based cohort study using an instrumental variable. *CMAJ*. 10 20 2015;187(15): 1125–1132. [PubMed: 26303244]
67. Yajnik CS, Chandak GR, Joglekar C, et al. Maternal homocysteine in pregnancy and offspring birthweight: epidemiological associations and Mendelian randomization analysis. *Int J Epidemiol*. 10 2014;43(5):1487–1497. [PubMed: 25052622]
68. Alwan NA, Lawlor DA, McArdle HJ, Greenwood DC, Cade JE. Exploring the relationship between maternal iron status and offspring's blood pressure and adiposity: a Mendelian randomization study. *Clin Epidemiol*. 2012;4:193–200. [PubMed: 22942651]
69. Lawlor DA, Timpson NJ, Harbord RM, et al. Exploring the developmental overnutrition hypothesis using parental-offspring associations and FTO as an instrumental variable. *PLoS Med*. 3 11 2008;5(3):e33. [PubMed: 18336062]
70. Wehby G, Jugessur A, Murray JC, Moreno L, Wilcox A, Lie RT. Genes as instruments for studying risk behavior effects: an application to maternal smoking and orofacial clefts. *Health Serv Outcome Res Methodol*. 7 1 2011;11(1–2):54–78.
71. Swanson SA, Hernandez-Diaz S, Palmsten K, Mogun H, Olfson M, Huybrechts KF. Methodological considerations in assessing the effectiveness of antidepressant medication continuation during pregnancy using administrative data. *Pharmacoepidemiol Drug Saf*. 9 2015;24(9):934–942. [PubMed: 26045370]
72. Kjaersgaard MI, Parner ET. Instrumental variable method for time-to-event data using a pseudo-observation approach. *Biometrics* 6 2016;72(2):463–472. [PubMed: 26574740]
73. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 1 15 2002;155(2):176–184. [PubMed: 11790682]
74. Collet JP, Schaubel D, Hanley J, Sharpe C, Boivin JF. Controlling confounding when studying large pharmacoepidemiologic databases: a case study of the two-stage sampling design. *Epidemiology* 5 1998;9(3):309–315. [PubMed: 9583424]

75. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol.* 8 1 2005;162(3):279–289. [PubMed: 15987725]
76. McDowell A, Engel A, Massey J, Maurer K. Plan and operation of the second National Health and Nutrition Examination Survey, 1976–80. National Center for Health Statistics. *Vital Health Stat.* 1981;1(15):1976.
77. National Health and Nutrition Examination Survey O. http://www.cdc.gov/nchs/data/nhanes/nhanes_13_14/2013-14_overview_brochure.pdf. 2013–2014.
78. Walker AM. *Observation and Inference An Introduction to the Methods of Epidemiology.* Newton Lower Falls: Epidemiology Resources Inc; 1991.
79. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology.* 7 2003;14(4):459–466. [PubMed: 12843772]
80. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf.* 5 2006;15(5):291–303. [PubMed: 16447304]
81. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med.* 8 15 2017;167(4):268–274. [PubMed: 28693043]
82. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol.* 12 1996;25(6):1107–1116. [PubMed: 9027513]
83. Khoury MJ, Flanders WD, James LM, Erickson JD. Human teratogens, prenatal mortality, and selection bias. *Am J Epidemiol.* 8 1989;130(2):361–370. [PubMed: 2750731]
84. Young JG, Tchetgen EJ, Hernan MA. The choice to define competing risk events as censoring events and implications for causal inference. 2018:arXiv:1806.06136
85. Young JG, Stensrud MJ, Didelez V, Robins JM, Hernan MA. Separable effects for causal inference in the presence of competing risks. 2019: arXiv:1901.09472.
86. MacDonald SC, Bateman BT, McElrath TF, Hernandez-Diaz S. Mortality and morbidity during delivery hospitalization among pregnant women with epilepsy in the United States. *JAMA Neurol.* Sep. 2015;72(9):981–988.
87. MacDonald SC, McElrath TF, Hernandez-Diaz S. Pregnancy outcomes in women with multiple sclerosis. *Am J Epidemiol.* 1 1 2019;188(1):57–66. [PubMed: 30165561]
88. Rothman KJ, Greenland S, Lash T. *Precision and Statistics in Epidemiologic Studies* In: Rothman K, Greenland S, Lash T, eds. *Modern Epidemiology.* 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008:148–167.
89. Lang J, Rothman K, Cann C. That confounded p-value. *Epidemiology.* 1998;9(1):7–8. [PubMed: 9430261]
90. Poole C. Beyond the confidence interval. *Am J Public Health.* 1987;77(2):195–199. [PubMed: 3799860]
91. Reefhuis J, Devine O, Friedman JM, Louik C, Honein MA, National Birth Defects Prevention Study. Specific SSRIs and birth defects: Bayesian analysis to interpret new data in the context of previous reports. *BMJ.* 7 8 2015;351:h3190. [PubMed: 26156519]
92. Andrade SE, Davis RL, Cheetham TC, et al. Medication exposure in pregnancy risk evaluation program. *Matern Child Health J.* 10 2012;16(7):1349–1354. [PubMed: 22002179]

KEY POINTS

- Preapproval clinical trials typically exclude pregnant women.
- Evidence on drug safety during pregnancy to inform drug labeling must come from postapproval controlled observational studies.
- Healthcare utilization databases are increasingly being used, to complement pregnancy registries and case-control studies.
- Healthcare utilization databases offer clear advantages in terms of size allowing to study rare exposure and outcomes, and rich information on potential confounders.
- There are important methodological considerations distinctive to perinatal epidemiologic studies that need to be attended to in order to generate valid, minimally biased risk.

FIGURE 1.

Impact of maternal and infant eligibility criteria on the size of the mother-infant linked Medicaid Analytic Extract (MAX) pregnancy cohort (2001-2007) [Colour figure can be viewed at wileyonlinelibrary.com]

Approach 1: Days' supply overlapping the etiologically relevant window*

- Advantages:
 - Maximizes capture of relevant exposures
- Disadvantages:
 - If women discontinue medication prior to end of days supply, will result in misclassification

Approach 2: ≥ 1 prescription dispensed during the etiologically relevant window

- Advantages:
 - Increases specificity for exposure during relevant window (women likely to take the medication immediately after dispensing)
- Disadvantages:
 - Loss of power

Approach 3: > 1 prescription dispensed during the etiologically relevant window

- Advantages:
 - Overcomes concern about adherence (if > 1 dispensing, medication likely being taken as prescribed)
 - Least vulnerable to misclassification of LMP
- Disadvantages:
 - Loss of power
 - Selection of unusual population

* The etiologically relevant window for congenital malformations is the first trimester of pregnancy, which is typically operationalized as the time period from LMP through LMP + 90 day.

FIGURE 2.

Approaches to operationalizing the definition of medication exposure using databases when studying the risks of congenital malformation [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 3.

Prevalence of major congenital malformations in the Medicaid Analytic Extract (MAX) defined on the basis of a single diagnostic claim or an algorithm requiring multiple codes or a single code plus a corrective procedure and/or infant death [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 4.

Adjusted relative risk of neonatal drug withdrawal according to maternal exposure to antipsychotic medications in addition to prescription opioids versus prescription opioids alone during the 45 days before delivery, according to level of adjustment for confounding. Medicaid Analytic Extract, 2000-2010. Partially adjusted: adjusted for specific opioid compound and dose; fully adjusted: adjusted for all predefined potential confounding variables. (Adapted from Huybrechts KF et al.⁵⁸ Risk of neonatal drug withdrawal after intrauterine co-exposure to opioids and psychotropic medications: cohort study. *BMJ* 2017;358:j3326) [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5.

External adjustment of the observed crude relative risk (RR_{ED}). Graph presents the adjustment of a hypothetical observed RR_{ED} of 1.50 (eg, between antidepressants and cardiac malformations) for smoking considering a range of associations between this confounder and the outcome (RR_{CD}) from 1 to 2, assuming an estimated prevalence of smoking of 22.0% among unexposed subjects, and a range of smoking (20%-70%) among exposed subjects

FIGURE 6.

Impact of restriction to live births showing the corrected relative risk as a function of the probability of live birth for infants with malformations among the unexposed, modeling a 10% and 20% lower probability of live birth among the exposed. Probability of live birth in nonmalformed fetuses is assumed to be 80%. Probability of live birth in malformed, unexposed fetuses is assumed to range from 75% to 55%. The potential impact of a 10% to 20% lower frequency of live birth among the medication exposed on relative risk estimates is estimated

TABLE 1

The positive predictive value of algorithms to define selected obstetrical outcomes in the Medicaid Analytic Extract based on medical chart review

Condition of Interest	Positive Predictive Value
Cardiac malformations	75.1 (64.4, 83.4)
Persistent pulmonary hypertension of the newborn ^a	89.6 (77.8, 95.5)
Multiparity	86.6 (79.9, 91.6)
Induction	96.6 (88.3, 99.1)
Cesarean delivery	98.7 (92.8, 99.8)
Preterm delivery	74.5 (60.5, 84.8)
Preeclampsia during a hospitalization ^a	94.5 (84.0, 98.3)

Note. Adapted from Palmsten K, Huybrechts KF, Kowal MK, Mogun H, Hernandez-Diaz S. Validity of maternal and infant outcomes within nationwide Medicaid data. *Pharmacoepidemiology and Drug Safety* 2014;23 (6):646-655.⁵

^aRestricting to those who were not transferred to another facility.

TABLE 2
Comparison of characteristics of different data sources for the study of drug safety in pregnancy

Pregnancy Registry	Electronic Health Records	Healthcare Utilization Database
Prospective data if enrolled before outcome	Prospective data recording	Prospective data recording
Ad hoc collection takes time and \$\$	Data exist (economy of cost and time)	Data exist (economy of cost and time)
Selected group of volunteers, limited follow-up	Real-world experience, stable population	Real-world experience, dynamic population
Information on one, or few, drug	Information on multiple drugs	Information on multiple drugs
Real use, outpatient and inpatient, Rx, and OTC	Outpatient prescriptions	Usually outpatient filling of prescription
Information on outcomes of interest	Information on multiple outcomes if recorded	Information on multiple outcomes if reimbursed
Incomplete ascertainment of pregnancy losses	Incomplete ascertainment of pregnancy losses	Incomplete ascertainment of pregnancy losses
Validation usually part of the design	May have access to validation	May have access to validation
Key clinical factors collected in detail	Broad range of clinical factors, with less granularity	Broad range of clinical factors, with less granularity
Can collect information on sociodemographics	Some information on sociodemographics	Little information on sociodemographics
May have laboratory data if collected	May have laboratory data	May have laboratory data in subsample
Can collect key factors (eg, gestational age, family history)	Key characteristics may be missing, eg, LMP (not always recorded), use algorithm	Key characteristics may be missing, eg, LMP, no claims for it, use algorithm
Some use external reference, few allow CER	Internal control groups allow CER	Internal control groups allow CER
Small populations	Large source population	Huge source population
Can target new drugs (need to recruit users)	No information on new drugs	No information on new drugs

Abbreviations: OTC, over the counter; LMP, last menstrual period.