Review

Longitudinal Methods for Modeling Exposures in Pharmacoepidemiologic Studies in Pregnancy

Mollie E. Wood, Angela Lupattelli, Kristin Palmsten, Gretchen Bandoli, Caroline Hurault-Delarue, Christine Damase-Michel, Christina D. Chambers, Hedvig M. E. Nordeng, and Marleen M. H. J. van Gelder*

* Correspondence to Dr. Marleen van Gelder, Department for Health Evidence, Radboud University Medical Center, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands (e-mail: Marleen.vanGelder@radboudumc.nl).

In many perinatal pharmacoepidemiologic studies, exposure to a medication is classified as "ever exposed" versus "never exposed" within each trimester or even over the entire pregnancy. This approach is often far from real-world exposure patterns, may lead to exposure misclassification, and does not to incorporate important aspects such as dosage, timing of exposure, and treatment duration. Alternative exposure modeling methods can better summarize complex, individual-level medication use trajectories or time-varying exposures from information on medication dosage, gestational timing of use, and frequency of use. We provide an overview of commonly used methods for more refined definitions of real-world exposure to medication use during pregnancy, focusing on the major strengths and limitations of the techniques, including the potential for method-specific biases. Unsupervised clustering methods, including *k*-means clustering, group-based trajectory models, and hierarchical cluster analysis, are of interest because they enable visual examination of medication use trajectories over time in pregnancy and complex individual-level exposures, as well as providing insight into comedication and drug-switching patterns. Analytical techniques for time-varying exposure methods, such as extended Cox models and Robins' generalized methods, are useful tools when medication exposure is not static during pregnancy. We propose that where appropriate, combining unsupervised clustering techniques with causal modeling approaches may be a powerful approach to understanding medication safety in pregnancy, and this framework can also be applied in other areas of epidemiology.

clustering methods; confounding factors (epidemiology); Cox models; epidemiologic methods; longitudinal studies; medication; pregnancy; time-varying exposure methods

INTRODUCTION

Studies of medication use in pregnancy present unique challenges when researchers need to ascertain exposure status. Pregnancy is time limited, not always planned, and frequently undetected in the first weeks or even months. Many outcomes of interest have a specific and narrow window of vulnerability to medication exposure (e.g., cardiac malformations occur as a result of exposures in gestational weeks 3–8) (1), whereas others have unknown or prolonged exposure vulnerability.

In longitudinal observational studies on medication use during pregnancy, valid and reliable exposure definitions are essential to prevent bias resulting from misclassification. In many pharmacoepidemiologic studies, exposure to a medication is classified as "ever exposed" versus "never exposed" during the pregnancy or within each trimester. This binary approach, however, does not reflect real-world exposure patterns; it does not distinguish between a single dose of medication and chronic use over many days, and it disregards important aspects such as dosage, treatment duration, and timing of exposure (2, 3). Consequently, the binary approach
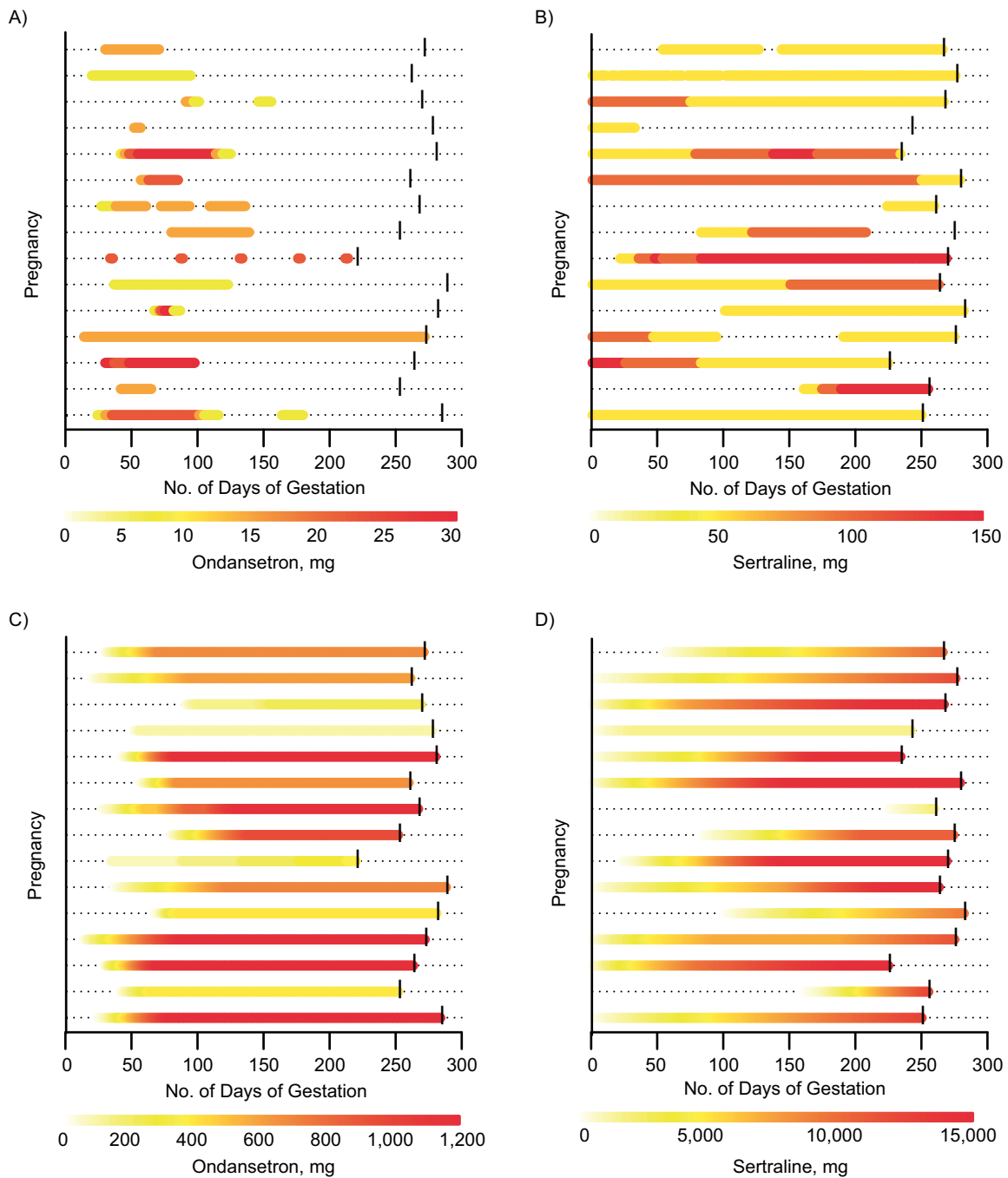
Figure 1. Heat maps showing daily (A, B) and cumulative dose (C, D) for ondansetron (A, C) and sertraline (B, D) use during pregnancy, with dose represented by gradations on the color spectrum from yellow (indicating lower daily and cumulative dose) to red (higher doses). Days without use of the medication of interest are denoted by horizontal dashes (A, B). Each of the 15 horizontal lines represents 1 pregnancy. The day of delivery is indicated by a vertical bar.

may lead to exposure misclassification for the vulnerable period of interest, because medication use may have taken place outside of sensitive time windows, even within the same trimester. A graphical presentation of this problem is given in Figure 1, in which the daily dose and cumulative dose of ondansetron, an antiemetic agent, and sertraline, an

antidepressant, on each day during pregnancy were plotted using a heat-map graphic for each individual (4). All the women depicted in Figure 1 would be classified as exposed, but clearly lumping all these individuals into a single "ever exposed" category during pregnancy for either of these medications does not accurately reflect the real-world

situation. In some studies, researchers have tried to address this issue by examining the cumulative days of medication use, and others have assessed dose–response categories (e.g., high, medium, or low dose vs. no exposure) of first or highest daily dose of medications used during the etiologically relevant gestational window (5–8). However, changes in medication exposure or dosage over time are not considered in these approaches (9).

Alternative exposure modeling methods can summarize complex, individual-level medication use trajectories or time-varying exposures from information on medication dose, gestational timing of use, and frequency of use into groups whose members have similar longitudinal exposure patterns. These medication exposure patterns may be used in many types of epidemiologic studies, including, but not limited to, drug utlization studies and exposure-outcome association analyses. Moreover, accurate exposure characterization is necessary in causal inference, because causal consistency, treatment variation irrelevance, and no measurement error are among the causal identification conditions (10). These methods are of particular interest for studies on safety of medication use during pregnancy, because the development of a specific perinatal outcome may depend on the dosage and duration of medication exposure within a specific gestational period.

To our knowledge, an overview of methods for more granular definitions of real-world exposure to medication during pregnancy is currently lacking; thus, our objective for this article was to review longitudinal methods for medication exposure modeling during pregnancy in epidemiologic studies. We outline 3 approaches to deal with complex longitudinal exposures: 1) unsupervised clustering methods as a mean to identify trajectories of medication use, 2) extended Cox models to correctly classify exposed and unexposed person-time, and 3) g-methods to explicitly model longitudinal exposures when confounding is affected by prior exposure. For each approach, we provide pregnancy study examples (Table 1) and elaborate on the strengths and limitations (Table 2), including the potential for biases associated with these approaches. Although we focus on perinatal studies, because of the importance of these methods for determining sensitive periods of development, these methods can also be applied in other areas of epidemiology, either separately or in combination. Software that can be used to apply the methods described is included in the Web Table 1 (available at https://doi.org/10.1093/epirev/mxab002).

UNSUPERVISED CLUSTERING METHODS

Unsupervised clustering methods are used to group individuals with similar patterns of values for a given variable or variables. The intent is to create homogenous groups that minimize within-group variance and maximize between-group variance. The methods are considered to be unsupervised because no a priori assumptions are made regarding group membership with respect to the outcome or other covariates. These methods may be used to identify groups with similar patterns of medication exposure (e.g., adherence, dose, and/or gestational timing of use) during preg-

nancy. Unsupervised clustering methods previously used in pregnancy medication studies, including k-means clustering, group-based trajectory models, and hierarchical cluster analysis, are described in the following sections. These methods can be used to identify clinically relevant patterns of real-world medication use. For example, Figure 2A shows k-means clustering–identified patterns of prednisone dose used by pregnant women with rheumatoid arthritis. In Figure 2A, group 1 includes women who took high doses of prednisone throughout pregnancy and group 2 contains women who increased their dose later in pregnancy, likely due to an increase in disease symptoms, whereas group 3 includes women with physiologic levels of prednisone throughout pregnancy.

k-Means clustering

k-Means is an unsupervised clustering method that has been used in studies with longitudinal data to identify similar patterns of values (i.e., trajectories), for 1 continuous variable or jointly for multiple, continuous, correlated variables measured at multiple points (11, 12). k-Means clustering is an algorithm for which the aim is to partition n observations (e.g., individuals) into k clusters. The method is nonparametric and no assumptions are made about the shape of the trajectories (11). Through a series of iterations, k-means minimizes the squared error between the cluster mean and points in the cluster for all clusters (13). Consequently, data points within the same cluster are considered to be more similar to each other, whereas data points in different clusters will be less similar. After initially assigning each observation to a cluster, k-means begins a 2-phase iterative algorithm to identify optimal clusters: 1) an expectation phase, in which the center of each cluster is calculated; and 2) a maximization phase, in which each observation is assigned to its nearest cluster (e.g., using Manhattan or Euclidean distance as the distance measure) (13). This process is repeated until there are no more changes in the clusters (convergence of the algorithm) or until the number of predefined iterations has been reached (13).

Quality criteria, such as the Caliński and Harabasz, Ray and Turi, and Davies and Bouldin criteria, can be used to help select the optimal number of clusters (13–15). However, using these criteria does not always result in convergence on a single solution, convergence on a clinically relevant solution, or identification of large-enough clusters to carry out additional analyses.

In previous studies, researchers have used k-means to identify patterns of psychotropic medication (e.g., antidepressants, anxiolytics, hypnotics), ondansetron, and prednisone use during and after pregnancy and to link the patterns with maternal and infant outcomes (Table 1; Figure 2A). The authors of these studies used data on medication exposure from prescription medication orders in patients' electronic health records (16–19), pharmacy dispensing information (9, 20, 21), or self-report (4, 22) to identify medication trajectories. In these studies, investigators linked trajectories of higher daily and/or cumulative doses as compared with trajectories of lower doses with perinatal outcomes. Trajectories of anxiolytic and hypnotic drug

Table 1. Overview of Studies on Medication Use During Pregnancy Using Longitudinal Methods for Exposure Modeling

First Author, Year (Reference No.)	Study Population	Exposure of Interest	Modeling Method	Outcome of Interest
Unsupervised Clustering Methods				
Hurault-Delarue, 2016 (9)	Women included in the EFEMERIS database who gave birth in Haute-Garonne, France, between 2004 and 2010	Prescriptions of psychotropic drugs, transformed into the number of DDDs per month	*k*-means clustering	None
Hurault-Delarue, 2017 (20)	Women included in the EFEMERIS database who delivered a liveborn infant in Haute-Garonne, France, between 2004 and 2010	Anxiolytic and hypnotic medications dispensed during pregnancy, transformed into the number of DDDs per month	*k*-means clustering	Neonatal pathology: oxygen therapy, intubation, resuscitation, transfer to specialized service, and/or respiratory distress
Bandoli, 2018 (17)	Pregnant women delivering at UC San Diego Health with ≥1 antidepressant prescription in the 3 months before or during pregnancy	Average daily dose and cumulative dose of antidepressants per week during the first 32 weeks of gestation and the first 13 weeks postpartum based on EMR	*k*-means clustering	Birth weight, gestational age at delivery
Palmsten, 2018 (4)	MotherToBaby Autoimmune Diseases in Pregnancy Study: pregnant women from the United States and Canada with rheumatoid arthritis and prednisone use	Daily and cumulative dose of prednisone during the first 32 weeks of gestation assessed with telephone interviews including start and stop dates, frequency of use, and strength	*k*-means clustering	Gestational age at delivery
Bandoli, 2020 (16)	Liveborn, singleton deliveries between 2012 and 2016 among girls and women aged 12–49 years identified in OptumLabs Data Warehouse administrative health-care claims	Antidepressant prescription fills between LMP and 35 gestational weeks with dosages converted to fluoxetine equivalents	*k*-means clustering	Major cardiac malformations, preterm birth, and newborn respiratory distress
Lemon, 2020 (18)	Liveborn, singleton deliveries at Magee-Womens Hospital with UPMC Health Plan coverage from 2006 through 2014	Ondansetron exposure extracted from the inpatient EMR and through insurance claims for outpatient prescriptions	*k*-means clustering	Neonatal cardiac anomalies
Palmsten, 2020 (21)	Liveborn deliveries between 2012 and 2016 among girls and women aged 12–49 years identified in OptumLabs Data Warehouse administrative health-care claims	Pharmacy dispensing of antidepressants from 3 months before LMP through 35 gestational weeks	*k*-means clustering	Preeclampsia and postpartum hemorrhage
Palmsten, 2020 (22)	MotherToBaby Pregnancy Studies: pregnant women from the United States and Canada with rheumatoid arthritis	Cumulative dose of oral corticosteroids during the first 139 days of gestation assessed with telephone interviews including start and stop dates and dose	*k*-means clustering	Preterm birth

Table continues

Table 1. Continued

First Author, Year (Reference No.)	Study Population	Exposure of Interest	Modeling Method	Outcome of Interest
Palmsten, 2021 (19)	Women with asthma or SLE enrolled in the California Medicaid program linked to birth certificates, 2007–2013	Outpatient pharmacy claims for oral corticosteroids and disease-related medications between LMP and gestational day 258	k-means clustering	Preterm birth
Frank, 2018 (28)	Pregnant women participating in MoBa using thyroid hormone replacement therapy	Daily doses of hypothyroid medication from 6 months prior to pregnancy until 12 months after delivery, based on prescriptions in NorPD (date of dispensing, strength, and quantity) and self-completed questionnaires	GBTM	None
Schaffer, 2019 (30)	Data linked for the MUMS Study: women who gave birth between 2005 and 2012 in New South Wales, Australia	Prescription for antipsychotics: total and average DDDs available in each 30-day interval during the study period	GBTM	Pregnancy complications and birth outcomes
Wood, 2021 (31)	Pregnancies enrolled in the IBM MarketScan health-care claims database between 2011 and 2015 resulting in a live or stillbirth	Outpatient claims for generic names of medications used in the treatment of migraine	GBTM and group-based multitrajectory models	None
Salvatore, 2017 (34)	Pregnant women participating in MoBa with paracetamol use	Questionnaire: any comedication used during pregnancy at 4-week intervals, including indication for use and number of days used	HCA	None
		Extended Cox Models		
Yonkers, 2011 (55)	Women <17 weeks of gestation from obstetrical practices and hospital-based clinics in Connecticut and western Massachusetts who underwent antidepressant treatment or had a current or prior history of a depressive disorder, between March 2005 and May 2009	Self-reported antidepressant use via structured at-home interview, asked to show pill bottles	Time-varying approach in Cox proportional hazard models	Major depressive episode
Xu, 2012 (52)	Vaccine and Medication in Pregnancy Surveillance System H1N1 Vaccine in Pregnancy Study: women enrolled before 20 weeks of gestation, US,	H1N1 vaccine	Time-varying approach in Cox proportional hazard models	Miscarriage
Matok, 2014 (47)	UK HES database linked to the CPRD: women between 15 and 45 years old who delivered a singleton live birth between April 1, 1997, and March 31, 2012	Decongestant prescriptions between gestational weeks 27–37 registered in CPRD	Time-varying approach (considered unexposed until prescription) in Cox proportional hazard models	Preterm birth

Table continues

Table 1. Continued

First Author, Year (Reference No.)	Study Population	Exposure of Interest	Modeling Method	Outcome of Interest
Daniel, 2015 (53, 54)	Pregnant women registered with the Clalit Health Services who were admitted for a delivery or had a miscarriage at Soroka Medical Center (Israel)	NSAIDs dispensed between LMP and the day before admission to the hospital for miscarriages or 20 weeks' gestation for pregnancies that ended with a birth	Time-varying approach (considered unexposed until prescription) in Cox proportional hazard models	Miscarriage
		G-Methods		
Bodnar, 2004 (66)	Iron Supplementation Study: women <20 weeks pregnant at the initial visit to a public prenatal clinic in Raleigh, North Carolina, 1997–1999	Randomly assigned to receive iron supplements; women were asked to return study pill bottles and to complete questionnaires on compliance. Pharmacy records on dispensing of iron-containing supplements	Marginal structural models	Anemia at delivery
Wood, 2016 (67)	Pregnant women participating in MoBa who had a singleton birth without major birth defects	Questionnaire: triptan use, with timing of exposure collapsed into trimester categories	Marginal structural models	Neurodevelopmental outcome at age 3 years
Lupattelli, 2017 (68)	Depressed pregnant women participating in MoBa	Questionnaire: antidepressant use during pregnancy, categorized in 4-week intervals	Marginal structural models	Preeclampsia
Lupattelli, 2018 (69)	Pregnant women participating in MoBa reporting depressive/anxiety disorders before and/or during pregnancy, linked to the Medical Birth Registry of Norway	Questionnaire: SSRI use at 4-week intervals during pregnancy, including indication for use and number of days used	Marginal structural models	Behavioral, emotional, and social development in preschool-aged children
Petersen, 2018 (70)	Pregnant women participating in the DNBC or MoBa	Paracetamol, aspirin, and ibuprofen DNBC: 3 telephone interviews, reported on a week-by-week basis MoBa: Questionnaire responses, reported in 4-week intervals	Marginal structural models	Cerebral palsy

Abbreviations: CPRD, Clinical Practice Research Datalink; DDD, defined daily dose; DNBC, Danish National Birth Cohort; EFEMERIS, Evaluation chez la Femme Enceinte des Medicaments et de Leurs Risques; EMR, electronic medical record; G-methods, generalized methods; GBTM, group-based trajectory model; HCA, hierarchical cluster analysis; HES, Hospital Episodes Statistics; LMP, last menstrual period; MoBa, Norwegian Mother and Child Cohort Study; MUMS, Maternal Use of Medications and Safety; NorPD, Norwegian Prescription Database; NSAID, nonsteroidal antiinflammatory drug; SLE, systemic lupus erythematosus; SSRI, selective serotonin reuptake inhibitor; UC, University of California; UPMC, University of Pittsburgh Medical Center.

Table 2. Summary of the Main Applications, Advantages, and Limitations of *k*-Means Longitudinal Clustering, Group-Based Trajectory Modeling, Hierarchical Cluster Analysis, Extended Cox Models, and G-Methods

Method	Specified by Researcher	Applicability	Advantage	Limitations
Unsupervised clustering methods				No guarantee that identified clusters are heuristically or clinically meaningful; vulnerability to biases (confounding, selection, misclassification) may be less apparent
k-means Longitudinal clustering	Number of clusters	Model similar patterns of values for longitudinally collected variables	Nonparametric; requires no assumptions about trajectory shape; optimizes an objective function (minimizing sum of squared error)	Assumptions of equal variances for *k* groups may not identify smaller groups; assumes clusters are linearly separable; will identify distinct groups in uniform data
Group-based trajectory models	Number and shape of trajectories; type of parametric model	Finite mixture model for assigning individuals to longitudinal trajectories, given similar values on variables of interest	Flexibility for handling different variable types (dichotomous, count, continuous)	Convergence problems when sample size is small or when specified trajectory numbers or shapes fit the data poorly
Hierarchical cluster analysis	Similarity definition; location of dendrogram cuts	Clusters observations based on researcher-defined values of similarity	Number of clusters not specified a priori; allows for flexible definitions by researcher	Computationally intense, may be infeasible in large data sets
Methods using a priori exposure definitions				A priori definitions for exposure may not capture the most common patterns or the clinically relevant window of vulnerability
Extended Cox models	Definition of exposure person-time, confounders, outcomes	Considers exposure as a function of time	Researcher can update exposure status during follow-up time; includes flexible considering of truncation and censoring	Cannot address cumulative, joint, or time-varying exposure with time-varying confounding
G-methods	Definition of exposure, outcomes, confounders	Scenarios where treatment and confounding changes over time	Model effect of time-varying treatment in the presence of feedback from time-varying confounding	Requires measurement of all relevant exposures and confounders over time

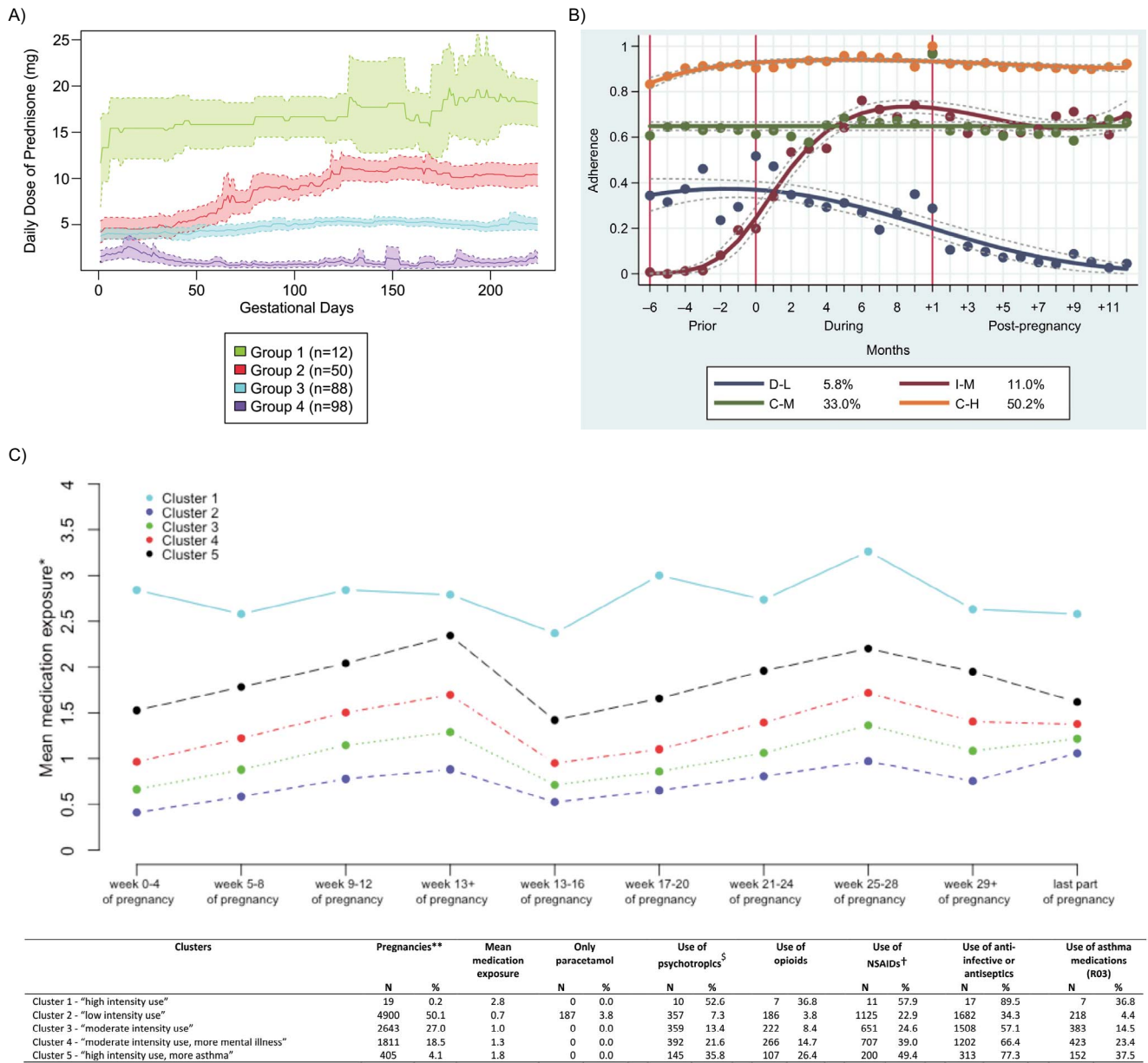Abbreviation: G-methods, Robins' generalized methods.

Figure 2. Data visualization of unsupervised clustering methods applied in pregnant women: A) *k*-means clustering of prednisone exposure among pregnant women with rheumatoid arthritis (4); B) group-based trajectory models of dispensed thyroid hormone replacement therapy (28); and C) hierarchical cluster analysis of the average number of medication exposures (34). ATC, Anatomical Therapeutic Chemical; C-H, constant-high; C-M, constant-medium; D-L, decreasing-low; I-M, increasing-medium.

dose identified by *k*-means were associated with differing neonatal risk, whereas binary exposed and unexposed groups had similar risks (20). Varying risks for postpartum hemorrhage were identified in another study on the basis of both gestational timing and dose of antidepressant exposure (21). Increased risks were observed for trajectories with

antidepressant exposures sustained later in pregnancy but not for a trajectory containing women with antidepressant dose reduction or discontinuation early in pregnancy. These examples illustrate that trajectory groups can be considered a possible method for defining exposure status in studies of medication safety during pregnancy.

Group-based trajectory modeling

Group-based trajectory modeling (GBTM) is an unsupervised method in which semiparametric models are used to identify longitudinal trajectories (23, 24). GBTM estimates multiple models simultaneously by maximizing a combined likelihood. Specifically, GBTM simultaneously estimates 1) a multinomial model for group-assignment probabilities and 2) models estimating longitudinal trajectories using polynomial functions of time. Individuals are assigned to the trajectory group to which they have the highest membership probability. Analysts can specify a range of number of groups and polynomial shapes of each group. Multiple models are compared using the Bayesian Information Criterion, the odds of correct classification, and other approaches, in combination with expert clinical opinion to select the optimal number of groups (25–27).

In the field of perinatal pharmacoepidemiology, at least 4 groups have used GBTM to study medication use (Table 1). Frank et al. (28) grouped women according to monthly probability of having used thyroid hormone replacement therapy before, during, and after pregnancy. GBTM identified 4 distinct patterns of use (Figure 2B), with lower education of the mother predicting membership in the lowest thyroid hormone replacement therapy use group. Other studies using GBTM grouped women according to the probability of filling an opioid prescription in each of 12 months after cesarean delivery (29) and antipsychotic use (30). In the latter study, women with the greatest exposure to antipsychotics had the highest rates of gestational hypertension and gestational diabetes. As with *k*-means, extensions of GBTM allow for simultaneous modeling of 2 or more exposures. Group-based multitrajectory modeling was recently used to model longitudinal fills of multiple prescription medications simultaneously: researchers examined polypharmacy patterns in pregnant women with migraine and compared maternal characteristics between the observed trajectories (31).

One drawback of GBTM is that, at times, it does not converge when nonparametric methods such as *k*-means do converge. In a study comparing the 2 methods using simulated data with known clusters, GBTM selected essentially the same trajectories as *k*-means in 3 data sets but did not converge in 1 data set, whereas *k*-means produced results consistent with known clusters (11). When performance was compared using data from 2 real cohort studies, results were again discrepant. In 1 data set, *k*-means and GBTM found trajectories that were quite similar. However, in the second data set, *k*-means resulted in 4 clusters, whereas GBTM either did not converge or gave incoherent results. The authors concluded that *k*-means seemed as efficient as the existing parametric algorithm when applied to polynomial data and potentially was more efficient when applied to nonpolynomial data (11).

Hierarchical cluster analysis

Hierarchical cluster analysis (HCA) is used to classify longitudinally measured characteristics into clusters on the basis of a customized distance measure informed by the researcher's prespecified definitions of similarity (32). This customized distance measure allows users to define "similarity" in the context of their research question (33). For modeling longitudinal medication exposures, women can be classified into clusters by use of different drugs over time. User-defined indices of similarity might include mechanism of action, indication for use, or even chemical structure of the active ingredient.

Similar to other approaches, the aim when using HCA is to identify homogenous groups within heterogeneous data. First, the possible features of medications are identified, and values are manually assigned by the researcher. Features might include the indication for use (e.g., analgesia vs. respiratory problems) or organ-system target (e.g., nervous vs. cardiovascular system). For example, if an analyst prioritizes indication as the feature of interest when considering concomitant medication use with paracetamol, opioids might be given a score of 1 and inhaled steroids a score of 3, indicating that opioids are more similar to paracetamol than inhaled steroids. Next, the distance between 2 observations, based on the totality of the features, is calculated. Clusters with the smallest distance between them are then merged. The merging is visually expressed using a dendrogram, where the height axis displays the distance between observations. Investigators then "cut" the dendrogram at clinically relevant levels, with the aim of identifying informative groups. In contrast to the other clustering methods discussed in this article (i.e., *k*-means and GBTM), the number of clusters is not specified as part of the HCA modeling process. Rather, solutions from different dendrogram cuts are compared and assessed for utility (32, 33).

HCA was used to capture longitudinal patterns of paracetamol use with concomitant medications during pregnancy (Table 1) (34). In their study, Salvatore et al. (34) used the difference between the codes of the Anatomical Therapeutic Chemical classification system, defining similarity between drugs as increasingly similar Anatomical Therapeutic Chemical codes (e.g., the code for paracetamol is N02BE01; that of ibuprofen in combination with codeine is N02AJ08; and the code for budesonide is R03BA02) used by the same individual. Paracetamol and ibuprofen plus codeine diverged at the third Anatomical Therapeutic Chemical level, whereas paracetamol and budesonide diverged at the first level, meaning paracetamol and ibuprofen plus codeine are more similar under this definition than are paracetamol and budesonide. Using this algorithm, Salvatore et al. (34) identified 5 clusters of medication users (Figure 2C). Two clusters were high-intensity users, differentiated by their use of medications for asthma; 2 clusters were moderate-intensity users, differentiated by their use of psychotropic drugs; the final cluster comprised low-intensity users. The flexibility of the similarity metric means that researchers can model 1 or multiple exposures simultaneously.

HCA offers several benefits for researchers: with no need to prespecify the number of possible groups, researchers can cut the dendrogram as they see fit to answer relevant research questions, taking into account practical considerations like sample size. The method can incorporate multiple variable forms, including categorical binary indicators, although it is not recommended to mix measurement scales (32). The

customized distance metric allows the analyst great flexibility to choose parameters best suited to the research question; however, HCA is computationally intensive compared with simpler methods like *k*-means and may not be best suited for larger data sets.

Challenges in interpreting results of unsupervised clustering methods

Unsupervised clustering methods (i.e., *k*-means clustering, GBTM, and HCA) can simplify dense medication exposure information while preserving some complexity regarding gestational timing of use and of dose changes, yielding more well-defined exposure groups than can the binary exposure approach. There are readily available software packages for implementing the methods (Web Table 1). However, several challenges must be addressed when using unsupervised clustering methods.

Identification versus overextraction of groups.   Medication exposure data are subject to multiple sources of error (35, 36), including under- or overreporting, lack of adherence, differences between dates of prescription fills and medication use, variability in doses, and inaccuracy of assigning the date of conception or start of pregnancy. If exposure data are recorded with error, unsupervised clustering methods may assign individuals to trajectory groups different from the ones that would have been assigned on the basis of actual use and could change the shape of the trajectories from what they would have been without errors. In addition, although clustering methods create more within-group homogeneity of the exposure(s) being modeled, there is still exposure heterogeneity within groups. Greater variability within clusters could potentially reduce the strength of exposure-outcome associations. Spaghetti plots of individual trajectories graphed against the group trajectory can help elucidate within-group heterogeneity (37).

Finally, methods used to identify unobserved groups or clusters are vulnerable to overextraction or identifying fictitious groups (38), and different clustering methods will likely produce different clusters (39). Applying unsupervised clustering methods does not necessarily result in useful or "true" clusters, and there is subjectivity in identifying the number of clusters for a study. In addition to considerations regarding group size, coupling clustering methods with clinical and biological knowledge is vitally important for identifying clinically relevant clusters of medication users.

Differences in gestational length.   The methods we have described thus far require exposure data during the entire exposure period of interest, and imputation methods are available if data are missing. However, in pregnancy research, exposure windows may differ between individuals because of different gestational lengths. For example, a woman who delivered at 34 gestational weeks would not have information on medication dose during pregnancy after 35 gestational weeks. In this example, it would not be sensible to impute the woman's medication use after delivery. To avoid imputation of medication use after delivery, some investigators have fit models only during

gestational weeks when all pregnancies were ongoing; for example, by excluding women delivering before 32 gestational weeks and focusing on medication exposures during the first 32 gestational weeks (4, 17).

Other challenges to consider.   With many unsupervised clustering methods available, there is no consensus on the most suitable clustering method given a certain data set (40). More research is needed to compare these methods in scenarios specific to pregnancy. Furthermore, it may be unclear whether modeling daily dose, cumulative dose, or another function of dose is best for a particular medication. An approach of modeling daily dose may be better suited for medications that are chronically used with relatively little variability, such as long-term use of a medication that some women discontinue or initiate during pregnancy (e.g., an antidepressant). In contrast, a monotonically increasing approach of modeling dose, such as cumulative dose, may be better suited for medications that change rapidly from a dose of 0 to a high dose and back to 0 within a few days (e.g., oral corticosteroids or opioids).

Challenges when using medication clusters for outcome estimation

Unsupervised clustering methods may be used in both descriptive drug utilization studies and exposure-outcome association analyses. In the latter case, it is important to carefully consider potential biases that could arise when estimating associations between medication trajectories and perinatal outcomes.

Bias from exposure misclassification.   In studies using a binary exposure definition, there is an expectation of bias towards the null if exposure misclassification is nondifferential with respect to the outcome; this is not necessarily the case with clustering methods if more than 2 groups best describe the data (41, 42). Studies may also use different methods to determine the date when pregnancy began; differences in these methods may lead to misclassification, especially for preterm births when using administrative claims-based algorithms (43).

Selection bias.   All unsupervised clustering approaches discussed may result in bias from selection due to differences in gestational length among the study participants and the need to avoid imputation of medication use after delivery. This may particularly affect studies that are focused on outcomes for which the etiologically relevant period includes late gestation. A central challenge in obtaining valid estimates in perinatal epidemiology is the omnipresent selection due to pregnancy losses occurring at every stage of gestation (44–46), and clustering methods are not immune to this. Furthermore, requiring pregnancies to progress to a certain point may reduce immortal time bias but increase selection bias.

Immortal time bias.   Classification of exposure during pregnancy can result in immortal time bias when entry into the exposed group is conditional on that pregnancy continuing long enough to have the opportunity for exposure, free of

the outcome of interest or any competing event. The time before exposure occurs is considered "immortal" because any outcome occurring before the opportunity for exposure would result in the event being assigned to the unexposed group, conferring an apparent protective effect of exposure (47–49). Pregnancies in which the outcome occurs before exposure would erroneously be classified as unexposed (48), although the pregnancies did not continue long enough to potentially become exposed to the medication. Pregnancies in which the outcome of interest is not experienced will thus have a greater chance (i.e., longer "survival") for being exposed to the medication during the follow-up.

Bias from confounding by indication. The core feature of unsupervised clustering methods is to group individuals together on the basis of similar longitudinal patterns of exposure. If researchers aim to investigate associations between trajectories and outcomes, the reasons for individuals to have similar trajectories of exposure must be evaluated. Consider hypothetical clusters of opioid use identified using GBTM in a cohort of patients who have migraine: a cluster going from intense use before pregnancy to low use through the first trimester versus a second cluster with consistently intense use before and during pregnancy. In this example, it is possible that the first cluster experienced remission of migraine with pregnancy onset, resulting in the trajectory of treatment in that cluster. If migraine status is a confounder that is not accounted for in the analysis, the resulting effect estimate may be biased. This problem is compounded when the potential confounder changes over time and is affected by past exposure (e.g., if migraine status changes over time and affects use of opioids at later time points in the pregnancy).

Although some clustering methods can incorporate time-varying covariates (50), these methods cannot be used to account for scenarios where time-varying confounding is affected by prior exposure. Time-varying confounding by underlying disease severity is especially a concern because changes in disease severity are often closely linked with changes in medication use and dosages of use and are often associated with the outcome. Thus, if trajectories or clusters of medication use are to be used in an outcome model (e.g., as the exposure in a model for risk of a perinatal outcome), researchers should consider whether methods to adjust for time-varying confounding are needed.

TIME-VARYING MEDICATION EXPOSURE AND PERINATAL OUTCOMES

Thus far, we have focused on modeling complex longitudinal exposures, and we have discussed the potential for bias with those approaches, including inability to adjust for time-varying confounders that also affect the exposure. In the following sections, we discuss 2 methods of estimating exposure parameters that address time-varying exposures.

The extended Cox model

Extended Cox models allow correct allocation of exposed and unexposed person-time during the follow-up (51); thus, they are useful in medication-in-pregnancy studies where the exposure status is a function of time. Like the Cox proportional hazard model, this method contains a baseline hazard function multiplied by an exponential function; however, in the extended Cox model, the exponential function contains both time-dependent and time-independent predictors (51). Time-dependent prenatal medication exposures are defined by an interaction term between the exposure variable and time t. The start of the follow-up time is usually set to the start of pregnancy.

A major assumption of the extended Cox model is that the hazard at time t depends on medication exposure status at the same time t and not on exposure status at later or earlier times (51). It is possible to allow lag-time variables for past medication exposure (51). Within this method, prenatal medication exposure status can be redefined during the follow-up time. A pregnancy is considered exposed only from the period of time after the actual intake of a medication; likewise, a pregnancy is considered unexposed from the beginning of the follow-up and up to the time of actual medication exposure, if exposure occurs (47). The correct allocation of time prevents immortal time bias (48).

Although the extended Cox model accounts for the exposure being a function of time, it provides a single regression coefficient for each time-varying exposure, which represents the overall estimate of the association between the time-dependent medication exposure and the perinatal outcome of interest (51). The interpretation of a resulting hazard ratio estimate would then be that at any given time t, the hazard for an unborn child who has already been exposed to a medication in utero is an estimated number of times higher than the hazard of an unborn child who has not been exposed to a medication by that time (but may be so later in gestation).

Some medication-in-pregnancy studies in which extended Cox models were applied exist in the literature (Table 1). In the majority of these associations were estimated between various prenatal medication exposures (i.e., nonsteroidal anti-inflammatory drugs, decongestants, or H1N1 vaccine) with proximal outcomes such as miscarriage and preterm birth (47, 52–54) or maternal major depressive relapse (55) (Table 1). Comparing the results of the standard versus extended Cox models shows that estimates from the time-independent model are biased because of immortal time (47, 52, 53), leading to an underestimation of increased risks or an overestimation of protective medication effects.

Extended Cox models can account for the exposure being a function of time, but there is no explicit modeling of time-varying exposure in relation to time-varying confounders. For instance, the extended Cox model would be unable to account for time-varying depression in pregnancy, which is affected by prior antidepressant exposure and determines future antidepressant exposure. The models are thus unable to provide time-specific estimates for the medication exposure (51), and their application does not overcome the potential bias introduced by time-varying confounding (56). For this, g-methods are required.

Generalized Methods

For many women, medication use is modified at pregnancy and throughout gestation, often in response to changes

in underlying disease status or perceived risk of exposure. For exposures that occur at a single time point, modeling approaches are well described in pregnancy literature (57) and in pharmacoepidemiology more generally (58). Exposures that change over time present a thornier methodological problem.

Consider an example of prenatal exposure to selective serotonin reuptake inhibitors (SSRIs), where the outcome of interest is risk of low birth weight in the infant. We may be interested in the effects of discontinuing SSRI use before the end of the first trimester, during mid-pregnancy, or continuing use throughout pregnancy. Even if the first SSRI treatment status were assigned randomly, an ethical study design would allow for treatment to change depending on the severity of maternal depressive symptoms: women with refractory depressive symptoms might have their dosages increased, whereas those responding well to cognitive behavioral therapy might decrease or discontinue SSRI treatment. In this example, depression severity is a time-varying confounder. To estimate the causal effect of SSRI exposure on offspring risk of low birth weight, we must control for depression severity. However, depression severity is affected by prior SSRI exposure, meaning that the current value for SSRI exposure has a causal effect on later depression severity. Depression severity is on the causal pathway between treatment and outcome and is simultaneously a confounder. Failing to adjust for depression severity will result in bias from residual confounding, but adjustment for depression severity blocks some of the effect of treatment and, in addition, could induce collider stratification bias (59) (Figure 3A and 3C).

In epidemiology, recognition of the problem of treatment-confounder feedback led to the development of the generalized (g) computation formula (g-formula) (56). Later innovations addressing similar problems under different assumptions included the inverse probability of treatment-weighted marginal structural models (IPTW-MSM) (60) and g-estimation of structural nested models (61). In non-parametric settings with no interaction between exposure and confounders, these methods will produce identical or near-identical results; however, in parametric settings, some differences will arise due to the assumptions and restrictions that are a function of the portion of the data being modeled.

Briefly, the IPTW-MSM approach estimates marginal effects with respect to time-varying confounders by specifying models for the treatment at each time point, conditional on treatment and confounder history. IPTW-MSM makes fewer assumptions than do the g-formula or g-estimation of structural nested models, and because there is no need to specify a model for the outcome or confounders, this approach is less prone to model misspecification. Weights can be inefficient or unstable, however, especially in high-dimensional settings where there are very rare or zero-probability combinations of exposure and confounders (62).

G-estimation of structural nested models takes a similar approach to IPTW-MSM but estimates effects conditional on time-varying confounders by explicitly modeling interactions between time-varying exposure and confounder (63). This approach is more efficient than IPTW-MSM, but very computationally intensive; it is particularly useful in settings

where effect modification is of interest. The g-formula estimates marginal effects by modeling the outcome conditional on confounders and treatment, and for each measurement of the time-varying confounder conditional on treatment and confounder history (62, 64). The g-formula method can easily handle complex joint interventions and is overall more efficient than IPTW-MSM and structural nested models if the model is correctly specified. However, these gains in complexity and efficiency come with several very strong assumptions. The g-null paradox means that the model is inconsistent with the causal null, whenever there is a non-null effect of past treatment on future covariates (62). It also requires modeling more of the data distribution, so misspecification is a larger concern than with other methods. Even in the absence of misspecification, the g-formula models potential outcomes under all possible interventions, so interpretation of results is difficult with high-dimensional treatments. These approaches allow for additional refinements, termed collectively "doubly robust" methods, which include targeted minimum loss-based estimation. Doubly robust methods model both the outcome and exposure mechanisms, and will yield unbiased effect estimates if either of these models is correctly specified (65). Targeted minimum loss-based estimation is a substitution estimator, meaning it is not prone to predicting values outside the sample, which is a drawback of singly robust g-methods (62).

Collectively, these methods offer the opportunity to consider hypothetical interventions that would change exposure from what was observed to prespecified regimes. Inherent in the g-methods approach is the design of a randomized controlled trial, either real or hypothetical. We can express the estimates from these models in terms of interventions: the effect of an intervention that causes everyone in our study to discontinue SSRI before the 13th week of gestation versus, for example, no intervention or natural course, or intervening to ensure everyone in our study continues SSRI use throughout pregnancy.

We return to the example of SSRI exposure and maternal depressive symptoms to illustrate the link between g-methods and the clustering methods we described previously in this article. In our hypothetical study, SSRI exposure is measured twice: at the 13th and 28th gestational weeks. At these measurements, participants are asked to report on their depressive symptoms, and their SSRI treatment is adjusted as needed. For the simplest binary-exposure definition, we would expect to see a group who used SSRIs at both times, a group with no SSRI use at any time, as well as groups that changed over the course of the study. In fact, there are 2^2, or 4, possible exposure groups, and if we allow 3 categories of SSRI exposure (none, low dose, high dose), there are 3^2, or 9, groups. Assuming that all pregnancies in our study proceed to term, we could fit a GBTM to the longitudinal SSRI exposure variable over the study period, with the goal of grouping together pregnancies with similar trajectories of treatment. The 4-group GBTM solution could include a group with constant high SSRI use, a group with initial SSRI use who discontinued in early pregnancy, and a group who discontinued early but resumed SSRI treatment during the latter part of gestation, and a group of never users. A generalization of these groups is illustrated in Figure 3B.
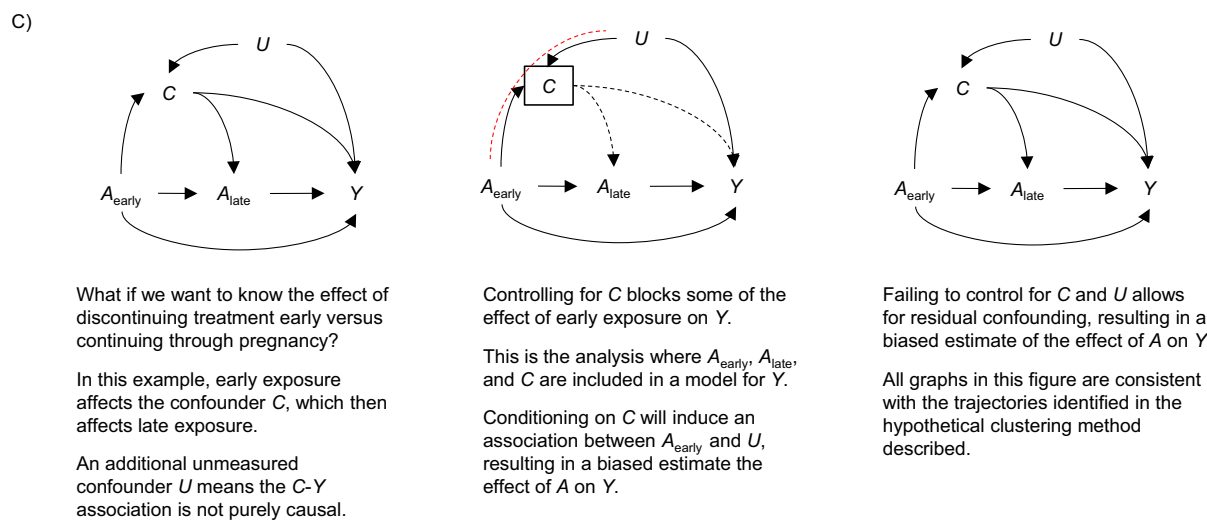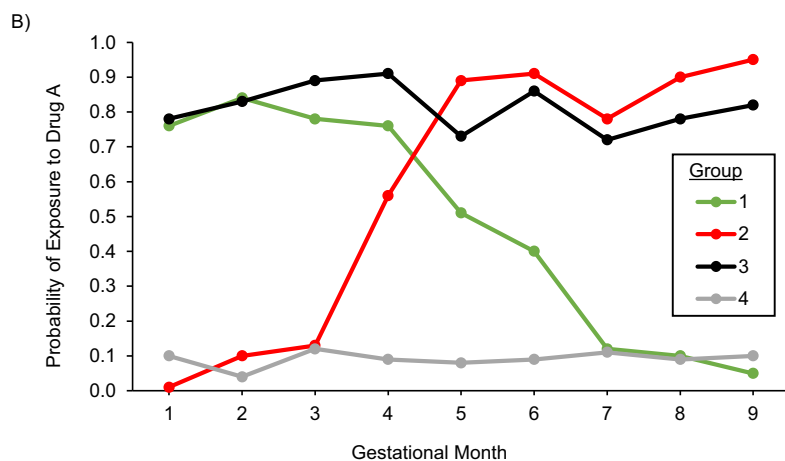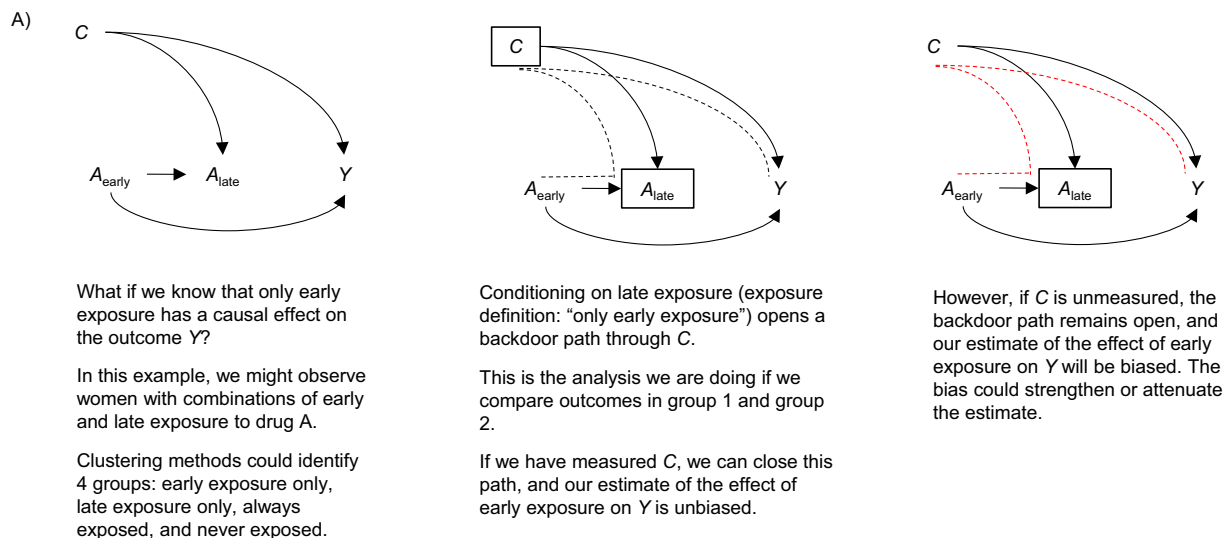
A)

What if we know that only early exposure has a causal effect on the outcome *Y*?

In this example, we might observe women with combinations of early and late exposure to drug A.

Clustering methods could identify 4 groups: early exposure only, late exposure only, always exposed, and never exposed.

Conditioning on late exposure (exposure definition: "only early exposure") opens a backdoor path through *C*.

This is the analysis we are doing if we compare outcomes in group 1 and group 2.

If we have measured *C*, we can close this path, and our estimate of the effect of early exposure on *Y* is unbiased.

However, if *C* is unmeasured, the backdoor path remains open, and our estimate of the effect of early exposure on *Y* will be biased. The bias could strengthen or attenuate the estimate.

B)

C)

What if we want to know the effect of discontinuing treatment early versus continuing through pregnancy?

In this example, early exposure affects the confounder *C*, which then affects late exposure.

An additional unmeasured confounder *U* means the *C*-*Y* association is not purely causal.

Controlling for *C* blocks some of the effect of early exposure on *Y*.

This is the analysis where A_{early}, A_{late}, and *C* are included in a model for *Y*.

Conditioning on *C* will induce an association between A_{early} and *U*, resulting in a biased estimate the effect of *A* on *Y*.

Failing to control for *C* and *U* allows for residual confounding, resulting in a biased estimate of the effect of *A* on *Y*.

All graphs in this figure are consistent with the trajectories identified in the hypothetical clustering method described.

Figure 3.   Using trajectories as exposures may mask time-varying confounding, resulting in biased estimates. A) Late exposure is a collider between early exposure and a confounder, *C*, opening a backdoor path to the outcome *Y*. C) Feedback between exposure and confounding means that estimates that do and do not control for *C* will be biased. A, C) Both causal models are consistent with trajectories described by the graph in (B). Group 1 is early exposure only; group 2 is late exposure only; group 3 is always exposed; and group 4 is never exposed.

The difficulty occurs when we use the groups defined by the GBTM in a model estimating the effect on low birth weight risk. Figure 3A and 3C illustrate different possible causal models that are both consistent with the trajectory groups, but in which the sensitive period of exposure and the confounding structure are very different. Although now masked by the trajectories, the problem of exposure-confounder feedback is still present: depressive symptoms are 1 of the causes of the different trajectory shapes; therefore, using the trajectories as our exposure implicitly adjusts for depressive symptoms. Estimates from this model will likely be biased. Although we can include time-varying covariates for depressive symptoms in our GBTM (50), this does not appropriately address treatment-confounder feedback.

However, g-methods and clustering approaches may still be complementary. For example, the g-formula is a model for all potential outcomes given all combinations of treatment and covariate history, making interpretation of results challenging when there are many variations of treatment, many measurements, or both. Using a clustering method to a priori identify treatment regimes of particular interest would mitigate this issue by allowing researchers to prioritize the most salient treatment patterns. This could be particularly effective in applications where the joint effects of polypharmacy are of interest.

IPTW-MSMs have had limited uptake in the pregnancy medication literature (Table 1). Examples include studies estimating the effect of iron supplementation with anemia at delivery (66), triptan exposure with neurodevelopment (67), antidepressant exposure with preeclampsia (68) and neurodevelopment (69), and paracetamol exposure with cerebral palsy (70). In these studies, researchers used marginal structural models to estimate effects of treatment at specific times in pregnancy, which could be used to identify higher risk-exposure windows. Using the marginal structural model method allowed for appropriate control of measured time-varying confounding from sources such as concomitant medication use or maternal depressive/anxiety symptoms. To our knowledge, other causal modeling approaches have not been applied to the study of medications in pregnancy.

Other complex problems in perinatal pharmacoepidemiology, including time-varying pregnancy losses, competing risks, and differences in gestational length, are addressable to some extent with g-methods, which can model time-varying censoring processes affected by treatment and covariates. Several g-methods tutorials are published and provide an accessible entry into a very dense literature (62–64, 71). Notably, these complex methods require well-measured, time-varying treatment and covariate data, which may not be readily available in the administrative databases often used to conduct pregnancy medication-safety studies (72) and may require augmentation with richer data sources (73). Biases described previously (e.g., identification of the start of pregnancy, misclassification) are still issues here; furthermore, even rich data sources often have information on disease severity collected a few times during pregnancy, but treatment can change many times, limiting the practical application of more complex methods.

IMPLICATIONS AND CLINICAL TRANSLATION

In this article, we have discussed a range of methods for addressing an important challenge in pregnancy medication research: how best to deal with complex longitudinal exposures. As larger data sources with increasingly granular medication-use data become more widely available, and as medication use among pregnant women increases (74–76), methods for modeling exposure must evolve in complexity to keep up. We have focused on presenting 2 distinct approaches for addressing complex longitudinal exposures. Using unsupervised clustering techniques allows researchers to conduct data-driven examinations of complex exposure patterns, which have been linked to perinatal outcomes. Causal modeling approaches, such as g-methods for estimating exposure effects in the presence of time-varying confounding, test the effect of prespecified exposure windows on outcomes. We suggest that combining these methods can be a powerful approach to understanding medication safety in pregnancy.

Unsupervised clustering methods are helpful for descriptive analyses because their use allows visual examination of medication use trajectories over time in gestation and complex individual-level exposures (4, 28, 31), and can provide insight into comedication and drug switching (31, 34), particularly when statistical results are combined with expert clinical opinion on the utility of the observed groups. Furthermore, unsupervised clustering methods can be useful for summarizing exposure from multiple time points in high-dimensional data such as from electronic health records (73). Researchers can use this approach to define and classify relevant windows of exposures as close as possible to real-world situations (2, 3), and the approach also can inform analytical studies. Using extended Cox models, analysts can redefine prenatal medication exposure status during follow-up (53), which limits the risk of immortal time bias, but extended Cox models cannot explicitly model joint or time-varying exposure in the presence of time-dependent confounders as g-methods do (56, 60, 77, 78). Construction of treatment episodes, time-varying confounders, cumulative exposure and latency, and treatment switching remain fundamental problems of time-varying methods, including extended Cox models and g-methods. G-methods, together with unsupervised clustering methods and descriptive analyses, can shed light into possible sensitive windows of medication exposure in pregnancy (67–69), which remain unknown for many perinatal and maternal outcomes (e.g., preterm birth, child neurodevelopment, preeclampsia).

We have described some of the more common approaches to addressing longitudinal exposure in medications-in-pregnancy studies, but other methods have been used as well. For instance, Bluhmki et al. (79) incorporated time-varying medication exposure during pregnancy and miscarriage as separate states in a multistate model to deal with the problem of left truncation and competing risks from other pregnancy outcomes. As this field continues to develop, it is critical that researchers evaluate methodologic novelty in terms of the potential gains but also the risks of bias. Development of a guideline for reporting results from unsupervised clustering methods, similar to existing guidelines for reporting results

from latent trajectory studies (80), would be a worthwhile endeavor.

Despite the advances and benefits of novel longitudinal exposure modeling methods, challenges remain. One such challenge is the lack of user-friendly quantitative bias analysis methods to correct for exposure misclassification that work with either clustering methods or time-varying exposures and confounders. Bias analysis is a useful tool to help researchers understand how much their effect estimates may be biased due to selection, misclassification, and confounding, and can incorporate both systematic and random error. These methods, however, largely assume a single binary exposure and outcome variable (81, 82), and their application to clustering methods for the exposure is unclear because of lack of methodological research on this topic. Importantly, the use of these more granular methods assumes that the quality of available data can support this kind of analysis. For example, the larger the discrepancy between medication prescription or dispensing information in electronic health record data and the actual dose and dates of medication use, the less useful these more complex methods are.

Another important challenge lies in how to communicate results from complex exposure analyses into clinical terms. Clinicians and women have so far interpreted risks according to trimester-specific exposures, dose, or as "ever" exposed in pregnancy, and so understanding more complex exposure patterns may be challenging. However, researchers may facilitate understanding by describing for each cluster the average daily dose and number of days of medication use during gestational windows.

In conclusion, longitudinal exposure methods are of particular interest in medication-in-pregnancy studies, because they can model complex exposures, shed light on potential vulnerable windows of exposure, and, ultimately, mirror real-world situations of medication use in pregnant women. Careful attention should be paid to the underlying assumptions, strengths and limitations, and potential for bias within each of these newer methods when conducting drug utilization or medication-safety studies in pregnancy. It is also essential to note that simpler binary approaches have some advantages over more complex methods, such as increasing power and minimizing some kinds of exposure misclassification. Efforts should be made to advance use of these newer methods in pregnancy research, where appropriate, and to maximize their utility in informing risks to maternal–child health.

ACKNOWLEDGMENTS

REFERENCES

1. Moore KL, Persaud TVN. *The Developing Human: Clinically Oriented Embryology*. 7th ed. Philadelphia, PA: Saunders; 2003.
2. Grzeskowiak LE, Gilbert AL, Morrison JL. Exposed or not exposed? Exploring exposure classification in studies using administrative data to investigate outcomes following medication use during pregnancy. *Eur J Clin Pharmacol.* 2012;68(5):459–467.
3. Pazzagli L, Linder M, Zhang M, et al. Methods for time-varying exposure related problems in pharmacoepidemiology: an overview. *Pharmacoepidemiol Drug Saf.* 2018;27(2):148–160.
4. Palmsten K, Rolland M, Hebert MF, et al. Patterns of prednisone use during pregnancy in women with rheumatoid arthritis: daily and cumulative dose. *Pharmacoepidemiol Drug Saf.* 2018;27(4):430–438.
5. Bérard A, Ramos É, Rey É, et al. First trimester exposure to paroxetine and risk of cardiac malformations in infants: the importance of dosage. *Birth Defects Res B Dev Reprod Toxicol.* 2007;80(1):18–27.
6. Mølgaard-Nielsen D, Pasternak B, Hviid A. Use of oral fluconazole during pregnancy and the risk of birth defects. *N Engl J Med.* 2013;369(9):830–839.
7. Palmsten K, Hernández-Díaz S, Huybrechts KF, et al. Use of antidepressants near delivery and risk of postpartum

hemorrhage: cohort study of low income women in the United States. *BMJ.* 2013;347:f4877.

8. Hernandez-Diaz S, Huybrechts KF, Desai RJ, et al. Topiramate use early in pregnancy and the risk of oral clefts. *Neurology.* 2018;90(4):e342–e351.

9. Hurault-Delarue C, Chouquet C, Savy N, et al. How to take into account exposure to drugs over time in pharmacoepidemiology studies of pregnant women? *Pharmacoepidemiol Drug Saf.* 2016;25(7):770–777.

10. Westreich D. Causal inference, causal effect estimation, and systematic error. In: Westreich D, ed. *Epidemiology By Design: A Causal Approach to the Health Sciences.* New York, NY: Oxford University Press; 2020:41–78.

11. Genolini C, Falissard B. Kml: a package to cluster longitudinal data. *Comput Methods Programs Biomed.* 2011; 104(3):E112–E121.

12. Genolini C, Alacoque X, Sentenac M, et al. Kml and kml3d: R packages to cluster longitudinal data. *J Stat Softw.* 2015; 65(4):1–34.

13. Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recogn Lett.* 2010;31(8):651–666.

14. Caliński T, Harabasz J. A dendrite method for cluster analysis. *Commun Stat.* 1974;3(1):1–27.

15. Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. *Psychometrika.* 1985;50(2):159–179.

16. Bandoli G, Chambers CD, Wells A, et al. Prenatal antidepressant use and risk of adverse neonatal outcomes. *Pediatrics.* 2020;146(1):e20192493.

17. Bandoli G, Kuo GM, Sugathan R, et al. Longitudinal trajectories of antidepressant use in pregnancy and the postnatal period. *Arch Womens Ment Health.* 2018;21(4): 411–419.

18. Lemon LS, Bodnar LM, Garrard W, et al. Ondansetron use in the first trimester of pregnancy and the risk of neonatal ventricular septal defect. *Int J Epidemiol.* 2020;49(2): 648–656.

19. Palmsten K, Bandoli G, Watkins J, et al. Oral corticosteroids and risk of preterm birth in the California Medicaid program. *J Allergy Clin Immunol Pract.* 2021;9(1):375–384.e5.

20. Hurault-Delarue C, Chouquet C, Savy N, et al. Interest of the trajectory method for the evaluation of outcomes after in utero drug exposure: example of anxiolytics and hypnotics. *Pharmacoepidemiol Drug Saf.* 2017;26(5):561–569.

21. Palmsten K, Chambers CD, Wells A, et al. Patterns of prenatal antidepressant exposure and risk of preeclampsia and postpartum haemorrhage. *Paediatr Perinat Epidemiol.* 2020; 34(5):597–606.

22. Palmsten K, Bandoli G, Vazquez-Benitez G, et al. Oral corticosteroid use during pregnancy and risk of preterm birth. *Rheumatology.* 2020;59(6):1262–1271.

23. Nagin D. *Group-Based Modeling of Development.* Cambridge, MA: Harvard University Press; 2005.

24. Franklin JM, Shrank WH, Pakes J, et al. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. *Med Care.* 2013; 51(9):789–796.

25. Nagin DS. Group-based trajectory modeling: an overview. *Ann Nutr Metab.* 2014;65(2–3):205–210.

26. Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociol Methods Res.* 2007;35(4):542–571.

27. Van Boven JFM, Koponen M, Lalic S, et al. Trajectory analyses of adherence patterns in a real-life moderate to severe asthma population. *J Allergy Clin Immunol Pract.* 2020;8(6):1961–1969.e6.

28. Frank AS, Lupattelli A, Matteson DS, et al. Maternal use of thyroid hormone replacement therapy before, during, and after pregnancy: agreement between self-report and prescription records and group-based trajectory modeling of prescription patterns. *Clin Epidemiol.* 2018;10:1801–1816.

29. Bateman BT, Franklin JM, Bykov K, et al. Persistent opioid use following cesarean delivery: patterns and predictors among opioid-naïve women. *Am J Obstet Gynecol.* 2016; 215(3):353.e1–353.e18.

30. Schaffer AL, Zoega H, Tran DT, et al. Trajectories of antipsychotic use before and during pregnancy and associated maternal and birth characteristics. *Aust N Z J Psychiatry.* 2019;53(12):1208–1221.

31. Wood ME, Burch RC, Hernandez-Diaz S. Polypharmacy and comorbidities during pregnancy in a cohort of women with migraine. *Cephalalgia.* 2021;41(3):392–403.

32. Kaufman L, Rousseeuw PJ. *Finding Groups in Data: An Introduction to Cluster Analysis.* Hoboken, NJ: John Wiley & Sons, Inc.; 2005.

33. Rokach L, Maimon O. Clustering Methods. In: Maimon O, Rokach L, eds. *Data Mining Knowledge and Discovery Handbook.* Boston, MA: Springer; 2005:321–352.

34. Salvatore S, Domanska D, Wood M, et al. Complex patterns of concomitant medication use: a study among Norwegian women using paracetamol during pregnancy. *PLoS One.* 2017;12(12):e0190101.

35. Van Gelder MMHJ, van Rooij IALM, de Walle HEK, et al. Maternal recall of prescription medication use during pregnancy using a paper-based questionnaire: a validation study in the Netherlands. *Drug Saf.* 2013;36(1):43–54.

36. Funk MJ, Landi S. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. *Curr Epidemiol Rep.* 2015;1(4):175–185.

37. Hickson RP, Annis IE, Killeya-Jones LA, et al. Opening the black box of the group-based trajectory modeling process to analyze medication adherence patterns: an example using real-world statin adherence data. *Pharmacoepidemiol Drug Saf.* 2020;29(3):357–362.

38. Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychol Methods.* 2003;8(3):338–363.

39. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. *J Clin Epidemiol.* 2012;65(10):1078–1087.

40. Rodriguez MZ, Comin CH, Casanova D, et al. Clustering algorithms: a comparative approach. *PLoS One.* 2019; 14(1):e0210236.

41. Walker A, Blettner M. Comparing imperfect measures of exposure. *Am J Epidemiol.* 1985;121(6):783–790.

42. Birkett NJ. Effect of nondifferential misclassification on estimates of odds ratios with multiple levels of exposure. *Am J Epidemiol.* 1992;136(3):356–362.

43. Zhu Y, Hampp C, Wang X, et al. Validation of algorithms to estimate gestational age at birth in the Medicaid analytic eXtract—quantifying the misclassification of maternal drug exposure during pregnancy. *Pharmacoepidemiol Drug Saf.* 2020;29(11):1414–1422.

44. Snowden JM, Bovbjerg ML, Dissanayake M, et al. The curse of the perinatal epidemiologist: inferring causation amidst selection. *Curr Epidemiol Rep.* 2018;5(4):379–387.

45. Liew Z, Olsen J, Cui X, et al. Bias from conditioning on live birth in pregnancy cohorts: an illustration based on neurodevelopment in children after prenatal exposure to organic pollutants. *Int J Epidemiol.* 2015;44(1):345–354.

46. Suarez EA, Landi SN, Conover MM, et al. Bias from restrict-ing to live births when estimating effects of prescription drug

use on pregnancy complications: a simulation. *Pharmacoepidemiol Drug Saf.* 2018;27(3):307–314.

47. Matok I, Azoulay L, Yin H, et al. Imortal time bias in observational studies of drug effects in pregnancy. *Birth Defects Res Part A Clin Mol Teratol.* 2014;100(9):658–662.

48. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol.* 2008;167(4):492–499.

49. Hutcheon JA, Savitz DA. Invited commentary: influenza, influenza immunization, and pregnancy—it's about time. *Am J Epidemiol.* 2016;184(3):187–191.

50. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol.* 2010;6(1):109–138.

51. Kleinbaum DG, Klein M. Extension of the Cox proportional hazards model for time-dependent variables. In: Kleinbaum DG, Klein M, eds. *Survival Analysis: A Self-Learning Text.* 3rd ed. New York, NY: Springer; 2012:241–288.

52. Xu R, Luo Y, Chambers C. Assessing the effect of vaccine on spontaneous abortion using time-dependent covariates Cox models. *Pharmacoepidemiol Drug Saf.* 2012;21(8):844–850.

53. Daniel S, Koren G, Lunenfeld E, et al. Immortal time bias in drug safety cohort studies: spontaneous abortion following nonsteroidal antiinflammatory drug exposure. *Am J Obstet Gynecol.* 2015;212(3):307.e1–307.e6.

54. Daniel S, Koren G, Lunenfeld E, et al. NSAIDs and spontaneous abortions - true effect or an indication bias? *Br J Clin Pharmacol.* 2015;80(4):750–754.

55. Yonkers KA, Gotman N, Smith MV, et al. Does antidepressant use attenuate the risk of a major depressive episode in pregnancy? *Epidemiology.* 2011;22(6):848–854.

56. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Math Modelling.* 1986;7(9–12):1393–1512.

57. Huybrechts KF, Bateman BT, Hernández-Díaz S. Use of real-world evidence from healthcare utilization data to evaluate drug safety during pregnancy. *Pharmacoepidemiol Drug Saf.* 2019;28(7):906–922.

58. Stürmer T, Wang T, Golightly YM, et al. Methodological considerations when analysing and interpreting real-world data. *Rheumatology.* 2020;59(1):14–25.

59. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004;15(5):615–625.

60. Robins JM, Hernan M, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11(5):550–560.

61. Robins JM, Blevins D, Ritter G, et al. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. *Epidemiology.* 1992;3(4):319–336.

62. Daniel RM, Cousens SN, De Stavola BL, et al. Methods for dealing with time-dependent confounding. *Stat Med.* 2013;32(9):1584–1618.

63. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol.* 2017;46(2):756–762.

64. Li X, Young JG, Toh S. Estimating effects of dynamic treatment strategies in pharmacoepidemiologic studies with time-varying confounding: a primer. *Curr Epidemiol Rep.* 2017;4(4):288–297.

65. Van der Laan MJ. Targeted maximum likelihood based causal inference: part I. *Int J Biostat.* 2010;6(2):2.

66. Bodnar L, Davidian M, Siega-Riz AM, et al. Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. *Am J Epidemiol.* 2004;159(10):926–934.

67. Wood ME, Lapane K, Frazier JA, et al. Prenatal triptan exposure and internalising and externalising behaviour problems in 3-year-old children: results from the Norwegian Mother and Child Cohort Study. *Paediatr Perinat Epidemiol.* 2016;30(2):190–200.

68. Lupattelli A, Wood M, Lapane K, et al. Risk of preeclampsia after gestational exposure to selective serotonin reuptake inhibitors and other antidepressants: a study from the Norwegian Mother and Child Cohort Study. *Pharmacoepidemiol Drug Saf.* 2017;26(10):1266–1276.

69. Lupattelli A, Wood M, Ystrom E, et al. Effect of time-dependent selective serotonin reuptake inhibitor antidepressants during pregnancy on behavioral, emotional, and social development in preschool-aged children. *J Am Acad Child Adolesc Psychiatry.* 2018;57(3):200–208.

70. Petersen TG, Liew Z, Nybo Andersen AM, et al. Use of paracetamol, ibuprofen or aspirin in pregnancy and risk of cerebral palsy in the child. *Int J Epidemiol.* 2018;47(1):121–130.

71. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol.* 2017;185(1):65–73.

72. Murray MD. Use of data from electronic health records for pharmacoepidemiology. *Curr Epidemiol Rep.* 2014;1(4):186–193.

73. Andrade SE, Bérard A, Nordeng HME, et al. Administrative claims data versus augmented pregnancy data for the study of pharmaceutical treatments in pregnancy. *Curr Epidemiol Rep.* 2017;4(2):106–116.

74. Andrade SE, Gurwitz JH, Davis RL, et al. Prescription drug use in pregnancy. *Am J Obstet Gynecol.* 2004;191(2):398–407.

75. Bjørn AMB, Nørgaard M, Hundborg HH, et al. Use of prescribed drugs among primiparous women: an 11-year population-based study in Denmark. *Clin Epidemiol.* 2011;3(1):149–156.

76. Smolina K, Hanley GE, Mintzes B, et al. Trends and determinants of prescription drug use during pregnancy and postpartum in British Columbia, 2002-2011: a population-based cohort study. *PLoS One.* 2015;10(5):e0128312.

77. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics.* 1992;48(2):479–495.

78. Gruber S, van der Laan MJ. An application of collaborative targeted maximum likelihood estimation in causal inference and genomics. *Int J Biostat.* 2010;6(1):18.

79. Bluhmki T, Fietz AK, Stegherr R, et al. Multistate methodology improves risk assessment under time-varying drug intake—a new view on pregnancy outcomes following coumarin exposure. *Pharmacoepidemiol Drug Saf.* 2019;28(5):616–624.

80. Van de Schoot R, Sijbrandij M, Winter SD, et al. The GRoLTS-checklist: guidelines for reporting on latent trajectory studies. *Struct Equ Model.* 2017;24(3):451–467.

81. Lash TL, Fox MP, MacLehose RF, et al. Good practices for quantitative bias analysis. *Int J Epidemiol.* 2014;43(6):1969–1985.

82. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data.* New York: Springer-Verlag; 2009.