Longitudinal Methods for Modeling Exposures in Pharmacoepidemiologic Studies in Pregnancy

Mollie E. Wood,[1†] Angela Lupattelli[2†], Kristin Palmsten,[3] Gretchen Bandoli,[4,5] Caroline Hurault-Delarue,[6] Christine Damase-Michel,[6] Christina D. Chambers,[4,5] Hedvig M.E. Nordeng[2,7] and Marleen M.H.J. van Gelder[8,9]*

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [2]PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway, [3]HealthPartners Institute, Minneapolis, MN, USA, [4]Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA, [5]Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA, [6]Service de Pharmacologie Médicale, Faculté de Médecine, Université Toulouse III, Centre Hospitalier Universitaire, INSERM1027, Toulouse, France, [7]Department of Child Development and Health, Norwegian Institute of Public Health, Oslo, Norway, [8]Department for Health Evidence, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, Netherlands and [9]Radboud REshape Innovation Center, Radboud university medical center, Nijmegen, Netherlands.

*Corresponding author. Department for Health Evidence (HP 133), Radboud university medical center, P.O. Box 9101, 6500 HB Nijmegen, Netherlands. E-mail: Marleen.vanGelder@radboudumc.nl

[†]Joint first authors.

Word count: 5,388 words

ABSTRACT

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 fails to incorporate important aspects such as dosage, timing of exposure, and treatment duration. Alternative exposure modeling methods can better summarize complex individual level medication utilization 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 as they allow for visual examination of medication utilization trajectories over time in pregnancy and complex individual-level exposures, as well as providing insight into co-medication and drug switching patterns. Analytical techniques for time-varying exposure methods, such as extended Cox models and g 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.

Key words:

Clustering methods; Confounding factors (Epidemiology); Cox Models; Epidemiologic methods; Longitudinal studies; Medication; Pharmacovigilance; Pregnancy; Time-varying exposure methods

KEY MESSAGES

- In longitudinal observational studies on medication use during pregnancy, exposure status is commonly simplified by classifying it into “ever exposed” versus “never exposed” over the gestation or within each trimester, which can elide important information on intensity and duration of exposure.
- Unsupervised clustering methods may be used to identify groups with similar patterns of medication exposure, for example based on adherence, dose, and/or gestational timing of use, during pregnancy.
- Extended Cox models are useful in studies in which the exposure status is a function over time, as they allow correct allocation of exposed and un-exposed person-time during follow-up.
- G methods offer the possibility to identify causal effects under time-varying exposure and confounding.
- Combining unsupervised clustering techniques with causal modeling approaches may improve studies of medication safety in pregnancy.

INTRODUCTION

Studies on 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), while others have unknown or prolonged exposure vulnerability (e.g., neurodevelopmental outcomes).

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" over the pregnancy, or within each trimester. This binary approach, however, does not reflect real-world exposure patterns, as it fails to distinguish between a single dose of medication and chronic use over many days: it disregards important aspects such as dosage, treatment duration, and precise timing of exposure (2, 3). Consequently, it may lead to exposure misclassification for the vulnerable period of interest, as 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 was plotted using a heat-map graphic for each individual (4). All of the women depicted in Figure 1 would be classified as exposed, but it is clear that lumping all these individuals into a single category "ever exposed" during pregnancy for either of these medications does not accurately reflect the real-world situation. Some studies have tried to address this issue by examining the cumulative days of medication use, and other studies have assessed dose-response categories (e.g., high, medium, or low dose versus 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 treatment intensity over time are not considered in these approaches (9).

Alternative exposure modeling methods can better summarize complex individual-level medication utilization trajectories or time-varying exposures from information on medication dosage, gestational timing of use, and frequency of use. These naturalistic medication exposure patterns may be used in many types of epidemiologic studies, including but not limited to drug utilization studies and exposure-outcome association analyses (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 intensity and duration of medication exposure within a specific gestational time period.

As an overview of methods for more refined definitions of real-world exposure to medication during pregnancy is currently lacking, the objective of this manuscript was to review longitudinal methods for medication exposure modeling during pregnancy in epidemiological studies. To this end, we outlined the methods commonly used to model time-varying exposures and longitudinal clustering methods, providing pregnancy study examples (Table 1) and elaborating on the strengths and limitations (Table 2), including the potential for biases associated with these methods. Although we will focus on perinatal studies due to the importance of these methods for determining sensitive periods of development, these methods can also be applied in other areas of epidemiology. Software for the methods described is included in the Supplementary Table.

LONGITUDINAL METHODS FOR MEDICATION EXPOSURE MODELING

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 pregnancy. Unsupervised clustering methods previously used in pregnancy medication studies, including *k*-means clustering, group-based trajectory models, and hierarchical cluster analysis, are described below.

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 for one continuous variable or jointly for multiple continuous correlated variables (28-30). *K*-means clustering is an algorithm that aims to partition *n* observations (e.g., individuals) into *k* clusters. The method is non-parametric and makes no assumptions about the shape of the trajectories (29). Through a series of iterations, *k*-means minimizes the squared error between the cluster mean and points in the cluster for all clusters (31). 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 two-phase iterative algorithm to identify optimal clusters: 1) expectation phase: the center of each cluster is calculated, and 2) maximization phase: each observation is assigned to its nearest cluster (using for example Manhattan or Euclidean distance as the distance measure) (29). This process is repeated until there are no further changes in the clusters (convergence of the algorithm) or until the number of predefined iterations has been reached (29).

Quality criteria, such as the Calinski and Harabasz, Ray and Turi, and Davies and Bouldin criteria, can be used to help select the optimal number of clusters (29, 32, 33). 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 further analyses.

Previous studies have used *k*-means to identify patterns of psychotropic medication use (e.g., antidepressants, anxiolytics and hypnotics), ondansetron and prednisone use during and after pregnancy and to link the patterns with infant outcomes (Table 1; Figure 2a). These studies used data on medication exposure from electronic health records (EHR) prescription medication orders (12, 13), pharmacy dispensing information (9, 11, 15), or self-report (4, 14) to identify medication trajectories. In these studies, investigators linked higher medication dose exposure profiles as compared with lower dose profiles with shorter gestational age at delivery and lower birth weight. They also found that *k*-means could identify different risks between clusters based on exposure intensity and cumulative dose, whereas comparisons between standard exposed and unexposed classified groups could not (4, 11, 12). These examples illustrate that trajectory groups can be considered as 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 that employs an underlying multinomial modeling strategy (34). Unlike *k*-means, GBTM methods use semiparametric models and provide formal statistical criteria for selecting the optimal number of groups. GBTM uses finite mixture models to gather individuals into similar trajectories. To do this, GBTM estimates multiple models simultaneously by maximizing a combined likelihood. With a specified probability distribution, GBTM assumes a set of polynomial functions of time can summarize individual differences in trajectories. Analysts specify the number of groups and the polynomial shape of each

group, and the results of multiple models are compared, using the Bayesian Information Criterion and the odds of correct classification, in combination with expert clinical opinion. (35).

In the field of perinatal pharmacoepidemiology, at least three groups have used GBTM to study medication utilization (Table 1). Frank *et al*. grouped women according to monthly probability of having thyroid hormone replacement therapy (THRT) before, during and after pregnancy (16). GBTM identified 4 distinct patterns of utilization (Figure 2b), with less maternal education predicting membership in the lowest THRT utilization group. Another study used GBTM to group women according to the probability of filling an opioid prescription in each of 12 months following cesarean delivery (36), whereas Schaffer *et al*. identified six trajectories of antipsychotic use (17). In the latter, women with the greatest exposure to antipsychotics had the highest rates of gestational hypertension and gestational diabetes. Although we did not identify any other pharmacoepidemiology studies in pregnancy that have applied GBTM, one study modeled both alcohol and cigarette consumption in pregnancy using GBTM, and assessed maternal characteristics of group assignment (37). Although this effort did not model joint trajectories (instead relying on separate trajectories of alcohol and cigarette use), as with *k*-means, it is possible to model joint trajectories in GBTM. Joint trajectory modelling will be useful in future perinatal pharmacoepidemiology work to assess concomitant exposures, such as antidepressants and benzodiazepines.

One drawback of GBTM is that, at times, it does not converge when non-parametric models such as *k*-means do converge. In a study comparing the two on simulated data with known clusters, *k*-means and GBTM performed similarly on 3 datasets, selecting essentially the same trajectories. On the fourth simulation, however, GBTM did not converge, while *k*-means produced results consistent with the known clusters (28). When performance was compared on data from two real cohort studies, results were again discrepant. In one dataset, *k*-means and GBTM found trajectories that

were quite similar. However, in the second real example dataset, *k*-means resulted in 4 clusters, while GBTM either failed to converge or gave incoherent results. The authors concluded that *k*-means seemed as efficient as the existing parametric algorithm on polynomial data, and potentially more efficient on non-polynomial data (28).

Hierarchical cluster analysis

Hierarchical cluster analysis (HCA) is a data reduction method that classifies longitudinally-measured characteristics into clusters based on a customized distance measure informed by the researcher's pre-specified definitions of similarity (38). For medication exposures, we would expect to classify users into clusters based on similarities between different drugs. This customized distance measure allows users to define "similarity" in the context of their research question (39). 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, HCA aims to identify homogenous groups within heterogeneous data. First, the possible features of medications are identified, and values are manually assigned. 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 two 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 other clustering methods discussed in this paper, such as GBTM, the number of clusters is not identified a

priori: rather, solutions from different dendrogram cuts are compared and assessed for utility (38, 39).

HCA was recently used to capture longitudinal patterns of paracetamol use with concomitant medications during pregnancy (Table 1) (10). The paper used the difference between Anatomical Therapeutic Chemical (ATC) codes, defining similarity between drugs as increasingly similar ATC codes (for example, paracetamol is N02BE01, ibuprofen in combination with codeine is N02AJ08, and budesonide is R03BA02). Paracetamol and ibuprofen-codeine diverge at the 3rd ATC level while paracetamol and budesonide diverge at the 1st level, meaning paracetamol and ibuprofen-codeine are more similar under this definition than paracetamol and budesonide. Using this algorithm, the study identified five clusters of paracetamol users (Figure 2c). Two clusters were high intensity users, differentiated by their use of medications for asthma; two clusters were moderate intensity users, differentiated by their use of psychotropic drugs; the final cluster comprised low intensity users.

HCA offers several benefits for researchers interested in data reduction: with no need to pre-specify the number of possible groups, researchers can cut the dendrogram as they see fit to answer relevant research questions. The method can incorporate multiple variable forms, including categorical binary indicators, although it is not recommended to mix measurement scales (38). The HCA method does not require data reduction prior to use, as can be the case with other methods like latent class analysis. The customized distance metric allows the analyst great flexibility to choose parameters best suited to the research question; however, HCA is computationally intensive compared to simpler methods like *k*-means, and may not be best suited for larger datasets.

Potential biases from unsupervised clustering methods

Unsupervised clustering methods 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 the traditional yes/no exposure approach. Furthermore, there are readily available software packages for implementing the methods (Supplemental Table). However, unsupervised clustering methods must be used with caution.

Bias from exposure misclassification

Medication exposure data can be subject to multiple sources of error (40), depending on the data source (e.g., self-report, administrative claims). Some examples of sources of exposure misclassification include: under- or over-reporting, 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, which could all result in misclassification of individuals into trajectory groups different from the ones that would fit actual trajectories of use. This misclassification could change the shape of the trajectories from what they would have been without errors (4). In addition, although clustering methods create more within-group homogeneity with respect to the exposure(s) being modeled, it is important to remember that there is still exposure heterogeneity within groups. Greater variability within clusters could potentially reduce the strength of exposure-outcome associations. Finally, studies using a simpler binary exposure definition benefit from the assumption that if exposure misclassification is nondifferential with respect to the outcome, there is an expectation of bias towards the null; this is not necessarily the case with clustering methods if more than two groups best describe the data (41, 42).

Bias from differences in gestational length

The methods described above 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, especially those overlapping with the third trimester, may differ between individuals because of different gestational lengths, and exposure data may be missing in an informative way. 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 useful to impute the woman's medication use after delivery for a study specifically looking at medication use during pregnancy. To avoid imputation of medication use after delivery, some investigators have fit models only during gestational weeks when all pregnancies were ongoing, e.g. by excluding women delivering before 32 gestational weeks and focusing on medication exposures during the first 32 gestational weeks (4, 12). All of these unsupervised clustering approaches may result in bias from selection; researchers should consider their specific question to determine which approach is most appropriate.

Other challenges to consider

Also, with many data reduction methods available, it may be unclear which method is best suited for a particular study. Further research to understand how these methods perform under different conditions is needed. 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 with some women discontinuing or initiating during pregnancy, e.g., antidepressant use. 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, triptans, or opioid use (4).

Importantly, applying unsupervised cluster selection methods does not necessarily result in useful or "true" clusters, and there is subjectivity in identifying the number of clusters for a study. Coupling clustering methods with clinical and biological knowledge is vitally important for identifying clinically relevant clusters of medication users.

Bias from confounding

The core feature of unsupervised clustering methods for exposure classification is to describe the tendency of a medication or medications to be associated with one another. In descriptive applications, this approach may be extremely useful for identifying similar patterns of use in complex data. However, if researchers aim to link use patterns to outcomes, reasons for this tendency must be considered. In the case of HCA, paracetamol and opioids might "tend" to be associated because they are both used to treat pain. If individuals use medications jointly for the same indication, and the indication is associated with the outcome, this can be a confounding issue. Similarly, compare two hypothetical clusters identified using GBTM: a cluster going from intense to low use, versus a second cluster with consistently low use. In this example, something happened to change the trajectory of the first cluster. If that "something" is a confounder that is not accounted for in the analysis, such as an improvement of the underlying indication, the resulting effect estimate may be biased.

To our knowledge, optimal methods to account for time-varying confounding in exposure profiles have not fully been addressed (43). Time-varying confounding by underlying disease severity is especially a concern as changes in disease severity (often linked with perinatal outcomes) are often closely linked with changes in medication use and dosages of use, and are associated with the outcome. Thus, methods to incorporate time varying confounding into longitudinal cluster analysis are needed.

Time-varying medication exposure within the extended Cox model

Extended Cox models allow correct allocation of exposed and unexposed person-time during the follow-up (21), and thus they are useful in longitudinal 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 (44). 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 date of last menstrual period or of conception.

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 (44). It is possible to allow lag-time variables for past medication exposure (44). Within this method, prenatal medication exposure status can be redefined during the follow-up time. A woman (and the unborn child) is considered as exposed only from the period of time following the actual intake of a medication; likewise, a woman (and the unborn child) is considered as unexposed from the beginning of the follow-up and up to the time of actual medication exposure (21).

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 (44). 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 applying extended Cox models exist in the current literature (Table 1). The majority were safety studies, and estimated associations of various prenatal medication exposures, i.e. non-steroidal anti-inflammatory drugs, decongestants, or H1N1 vaccine, with immediate birth outcomes such as miscarriage and prematurity (19-22). One effectiveness study also applied the method (18): Yonkers *et al*. examined the risk of maternal major depressive relapse

according to treatment with antidepressants, and allowed for lag-time exposure variables (Table 1). All safety studies replicated their analyses using an additional time-fixed analysis (19-21), which has the advantage of addressing the often overlooked concern of immortal time bias (45). In brief, this type of bias can arise if the time between the start of pregnancy (or of a specific time window) until initiation of medication, is misclassified as exposed. Indeed, a pregnancy would be truly exposed only from the actual time of treatment initiation and forward (45). Pregnancies in which the outcome occurs before exposure would erroneously be classified as unexposed (45), although they did not 'survive' long enough to potentially become exposed to the medication. Pregnancies that do not experience the outcome of interest will thus have greater chance (longer 'survival') for being exposed to the medication during the follow-up. As shown in the cited safety studies (19-21), and thoroughly described by Suissa (45), immortal time can lead to unexpected findings, which are often an underestimation of increased risks or an overestimation of protective medication effects.

Biases in extended Cox model applications

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. The models are thus unable to provide time-specific estimates for the medication exposure (44), and their application does not overcome the potential bias introduced by time-varying confounding (46). For this, g methods are required. Thus, bias due to time-varying confounding remains a concern with this method.

The extended Cox model is not immune from risk of bias due to misclassification of exposure. Because the exposure status is a function of time, accurate information on the date of medication treatment initiation is crucial to minimize this risk of bias. Ahrens *et al*. (47) has shown that a quantitative correction for exposure misclassification can be implemented within the extended Cox framework. In this work on the association between influenza vaccination in pregnancy and risk of

preterm birth, the authors applied probabilistic bias analysis that incorporated the date of vaccine assignment. Using bias parameters from internal and external validation data, the resulting point estimates corrected for exposure misclassification were slightly higher and less precise than those produced in the time-fixed analysis.

The extended Cox model is one of several available methods to minimize the risk of immortal time bias in longitudinal pregnancy studies. However, the method is vulnerable to other sources of bias just like more traditional methods are, including unmeasured and residual confounding, selection bias and attrition, and bias due to outcome misclassification.

G methods

In the situation where medications are used at a single time point, all confounders are measured, adherence is perfect, and there is no feedback between exposure assignment and post-baseline confounding, estimation of causal effects of treatment is relatively straightforward. However, these conditions rarely hold, especially for medications taken over weeks or months. *G* methods, which include inverse-probability weighted marginal structural models (MSM), structural nested models, and the g formula, offer the possibility of identifying causal effects in applications where feedback exists between time-varying exposures and confounders (48-51).

In pregnancy exposure research, investigators may be interested in four effects: the effect of never being exposed, the effect of being exposed only early in the pregnancy, the effect of being exposed only later in the pregnancy, and the effect of being exposed both early and late. Traditional methods make it relatively straightforward to estimate the effect of late exposure, with and without early exposure, by simply fitting a model for late exposure while controlling for early exposure. However, using standard regression techniques to estimate the effect of early exposure while

controlling for later exposure may return biased estimates as a result of conditioning on future events (Figure 3). Late exposure is on the causal pathway between early exposure and the outcome, and conditioning on it may either attenuate the true effect (if no late exposure-outcome confounding is present) or result in unpredictable bias due to conditioning on a collider in the presence of late exposure-outcome confounding.

A fundamental difference between g methods and unsupervised clustering methods is that g methods are not data-adaptive and do not assign exposure groups or status based on characteristics of the data. Rather, g methods rely on investigators to identify treatment or exposure regimes *a priori* and to model these regimes. For dichotomous exposures measured at a single time point, this is straightforward. However, as demonstrated in the pregnancy medication literature, exposure patterns during pregnancy are often complex (2), with many women discontinuing or switching their medications, as well as adding or removing concomitant medications. If we think of clustering methods such as *k*-means and GBTM as methods used to identify patterns over time, g methods can help investigators correctly model effects on outcomes for these trajectories.

Several examples of inverse probability of treatment weighted MSMs exist in the pregnancy medication literature (Table 1), including studies estimating the effect of iron supplementation on anemia at delivery (23), triptan exposure on neurodevelopment (24), antidepressant exposure on preeclampsia and neurodevelopment (25, 26), and paracetamol exposure on cerebral palsy (27). These studies use MSMs to estimate effects of treatment at specific times in pregnancy, and the MSM method allows for appropriate control of measured time-varying confounding, such as concomitant medication use or maternal depressive/anxiety symptoms. To the best of our knowledge, g estimation of structural nested models and the g formula have not been used in the pregnancy medication safety literature. For a more comprehensive introduction to g methods, we suggest work from Naimi *et al*. (51).

A major strength of g methods is their ability to explicitly model time-varying exposure, including feedback between exposures and confounders. As part of the larger field of causal inference, g methods clearly set out the assumptions necessary to interpret effect estimates as being causal. In pregnancy medication safety studies, we are interested in causal effects of medications (52, 53); delineating the conditions under which we can interpret estimates as causal effects, not simply associations, is a strength of this approach. In addition, methods can accommodate censoring as a function of time-varying exposure and confounding, which is potentially important for studies considering pregnancy loss.

Biases in g methods application

Causal modeling, and specifically g methods, have focused on bias from time-varying confounding as the central threat to validity in observational research. Importantly, g methods provide a way to adjust for observed confounders, but any unmeasured confounding may still result in bias, which can be towards or away from the null. Bias from selection into the study (e.g., only including pregnancies that have advanced to 22 weeks or more) or out of the study (e.g., including only live births or children followed up to a certain age) can be mitigated if it has been measured, using censoring weights, but can otherwise be a serious and underappreciated problem for g methods as well as more traditional approaches. Other major sources of bias, such as measurement error, have received limited attention, and g methods are vulnerable to bias from exposure and outcome misclassification in the same way as all previously described methods.

IMPLICATIONS AND CLINICAL TRANSLATION

In this paper, we discussed a range of methods for addressing an important challenge in pregnancy medication research: how best to deal with complicated longitudinal exposures. As larger data sources with increasingly granular medication use data become more widely available, and as medication use among pregnant women increases (54, 55), methods for modeling exposure must evolve in complexity to keep up. We have focused on presenting two distinct approaches for addressing complex longitudinal exposures. Unsupervised clustering techniques allow 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 pre-specified exposure windows on outcomes. Rather than seeing these approaches as opposed or mutually exclusive, 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 as they allow visual examination of medication utilization trajectories over time in gestation and complex individual-level exposures (4, 16), and can provide insight into co-medication and drug switching (10), particularly when statistical results are combined with expert clinical opinion as to the utility of the observed groups. This can aid researchers in defining and classifying relevant windows of exposures as close as possible to real-world situations (2, 3), and also inform analytical studies. Analytic techniques for time-varying exposure methods are a useful tool when a medication exposure is not static during pregnancy. Extended Cox models allow analysts to redefine prenatal medication exposure status during follow-up time (21), which limits the risk of immortal time bias, but cannot explicitly model joint or time-varying exposure in the presence of time-dependent confounders as g methods do (48-51). Construction of treatment episodes, time-varying confounders, cumulative exposure and latency, and treatment switching remain fundamental problems of time-varying methods (3). G

methods together with unsupervised clustering methods and descriptive analyses can shed light into possible sensitive windows of medication exposure in pregnancy (24-26), which remain unknown for many perinatal and maternal outcomes (e.g., preterm birth, child cognitive and behavioral development, or preeclampsia).

We 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*. (56) 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 (57), would be a worthwhile endeavor.

Despite the advances and benefits of novel longitudinal exposure modelling 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 (58) and so their application and suitability to clustering methods for the exposure is unclear due to 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 secondary health care data and the actual dose and dates

of medication use, the less useful these more complex methods may be when applied to such data sources.

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. Results of clustering methods can enable a better understanding of higher or lower health risks by longitudinal pattern of exposure, which is crucial for clinical decision-making and regulatory labeling.

In conclusion, longitudinal exposure methods are of particular interest in medication in pregnancy studies, as 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.

FUNDING

This work was supported by the International Society for Pharmacoepidemiology; the National Heart, Lung and Blood Institute [grant number T32HL098048-11 to M.E.W]; the Research Council of Norway [FRIMEDBIO grant number 288696 to A.L.]; and the European Research Council Starting Grant "DrugsInPregnancy" [grant number 639377 to H.M.E.N.]. K.P. was supported by a career development award from the Eunice Kennedy Shriver National Institute of Child Health & Human Development, National Institutes of Health [grant number R00HD082412].

REFERENCES

1. Moore KL, Persaud TVN. The developing human: clinically oriented embryology. 7th ed. Philadelphia: 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; 68:459–67.
3. Pazzagli L, Linder M, Zhang M, et al. Methods for time-varying exposure related problems in pharmacoepidemiology: an overview. Pharmacoepidemiol Drug Saf 2018; 27:148–60.
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; 27:430–38.
5. Bérard A, Ramos É, Rey É, Blais L, St.-André M, Oraichi D. First trimester exposure to paroxetine and risk of cardiac malformations in infants: the importance of dosage. Birth Defects Res B Dev Reprod Toxicol 2007; 80:18–27.
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; 369:830–39.

7. Palmsten K, Hernández-Diaz 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:e342–351.

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; 25:770–77.

10. Salvatore S, Domanska D, Wood M, Nordeng H, Sandve GK. Complex patterns of concomitant medication use: a study among Norwegian women using paracetamol during pregnancy. PLoS One 2017; 12:e0190101.

11. 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; 26:561–69.

12. Bandoli G, Kuo GM, Sugathan R, Chambers CD, Rolland M, Palmsten K. Longitudinal trajectories of antidepressant use in pregnancy and the postnatal period. Arch Womens Ment Health 2018; 21:411–19.

13. 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. Advance Access: December 20, 2019. DOI: 10.1093/ije/dyz255.

14. Palmsten K, Bandoli G, Vazquez-Benitez G, et al. Oral corticosteroid use during pregnancy and risk of preterm birth. Rheumatology. Advance Access: September 30, 2019. DOI: 10.1093/rheumatology/kez405.

15. Palmsten K, Chambers CD, Wells A, Bandoli G. Patterns of prenatal antidepressant exposure and risk of preeclampsia and postpartum haemorrhage. Paediatr Perinat Epidemiol. Advance Access: March 24, 2020. DOI: 10.1111/ppe.12660.

16. Frank AS, Lupattelli A, Matteson DS, Nordeng H. 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; 10:1801–16.

17. Schaffer AL, Zoega H, Tran DT, Buckley NA, Pearson S, Havard A. Trajectories of antipsychotic use before and during pregnancy and associated maternal and birth characteristics. Aust N Z J Psychiatry 2019; 53:1208–21.

18. Yonkers KA, Gotman N, Smith MV, et al. Does antidepressant use attenuate the risk of a major depressive episode in pregnancy? Epidemiology 2011; 22:848–54.

19. Xu R, Luo Y, Chambers C. Assessing the effect of vaccine on spontaneous abortion using time-dependent covariates Cox models. Pharmacoepidemiol Drug Saf 2012; 21:844–50.

20. Matok I, Azoulay L, Yin H, Suissa S. Immortal time bias in observational studies of drug effects in pregnancy. Birth Defects Res A Clin Mol Teratol 2014; 100:658–62.

21. Daniel S, Koren G, Lunenfeld E, Levy A. Immortal time bias in drug safety cohort studies: spontaneous abortion following nonsteroidal antiinflammatory drug exposure. Am J Obstet Gynecol 2015; 212:307.e1–e6.

22. Daniel S, Koren G, Lunenfeld E, Levy A. NSAIDs and spontaneous abortions - true effect or an indication bias? Br J Clin Pharmacol 2015; 80:750–54.

23. Bodnar LM, Davidian M, Siega-Riz AM, Tsiatis AA. Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. Am J Epidemiol 2004; 159:926–34.

24. Wood ME, Lapane K, Frazier JA, Ystrom E, Mick EO, Nordeng H. 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; 30:190–200.

25. Lupattelli A, Wood M, Lapane K, Spigset O, Nordeng H. Risk of preeclampsia after gestational exposure to selective serotonin reuptake inhibitors and other antidepressants: a study from

The Norwegian Mother and Child Cohort Study. Pharmacoepidemiol Drug Saf 2017; 26:1266–76.

26. Lupattelli A, Wood M, Ystrom E, Skurtveit S, Handal M, Nordeng H. 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; 57:200–08.
27. 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; 47:121–30.
28. Genolini C, Falissard B. KmL: k-means for longitudinal data. Comput Stat 2010; 25:317-328.
29. Genolini C, Falissard B. Kml: a package to cluster longitudinal data. Comput Meth Prog Bio 2011; 104:E112–21.
30. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R packages to cluster longitudinal data. J Stat Softw 2015; 65:1–34.
31. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recogn Lett 2010; 31:651–66.
32. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat 1974; 3:1–27.
33. Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. Psychometrika 1985;50:159–79.
34. Nagin DS. Group-based trajectory modeling: an overview. Ann Nutr Metab 2014; 65:205–10.
35. Nagin DS, Jones BL, Passos VL, Tremblay RE. Group-based multi-trajectory modeling. Stat Methods Med Res 2018; 27:2015–23.
36. Bateman BT, Franklin JM, Bykov K, et al. Persistent opioid use following cesarean delivery: patterns and predictors among opioid-naive women. Am J Obstet Gynecol 2016; 215:353.e1–18.
37. Dukes K, Tripp T, Willinger M, et al. Drinking and smoking patterns during pregnancy: development of group-based trajectories in the Safe Passage Study. Alcohol 2017; 62:49–60.

38. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. Hoboken: John Wiley & Sons, Inc., 2005.

39. Rokach L, Maimon O. Clustering methods. In: Maimon O, Rokach L. Data mining knowledge and discovery handbook. Boston: Springer, 2005; 321–52.

40. Van Gelder MMHJ, van Rooij IALM, de Walle HEK, Roeleveld N, Bakker MK. Maternal recall of prescription medication use during pregnancy using a paper-based questionnaire: a validation study in the Netherlands. Drug Saf 2013; 36:43–54.

41. Walker AM, Blettner M. Comparing imperfect measures of exposure. Am J Epidemiol 1985; 121:783–90.

42. Birkett NJ. Effect of nondifferential misclassification on estimates of odds ratios with multiple levels of exposure. Am J Epidemiol 1992; 136:356–62.

43. Jones BL. Traj group-based modeling of longitudinal data. Example (strtxmpl4.sas): creating start values to fit a trajectory model with time-varying covariates (TCOV). Available from: https://www.andrew.cmu.edu/user/bjones/strtxmpl4.htm (3 April 2020, date last accessed).

44. Kleinbaum DG, Klein M. Extension of the Cox proportional hazard model for time-dependent variables. In: Kleinbaum DG, Klein M. Survival analysis: a self-learning text. 3rd ed. New York: Springer, 2012; 241–88.

45. Suissa S. Immortal time bias in pharmacoepidemiology. Am J Epidemiol 2008; 167:492–99.

46. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11:550–60.

47. Ahrens K, Lash TL, Louik C, Mitchell AA, Werler MM. Correcting for exposure misclassification using survival analysis with a time-varying exposure. Ann Epidemiol 2012; 22:799–806.

48. 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; 7:1393–1512.

49. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. Biometrics 1992;48:479–95.

50. Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. Longitudinal data analysis. Boca Raton: Taylor & Francis Group LLC, 2009: 553–99.

51. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. Int J Epidemiol 2017; 46:756–62.

52. Hernán MA. Does water kill? A call for less casual causal inferences. Ann Epidemiol 2016; 26:674–80.

53. Wood ME, Lapane KL, van Gelder MMHJ, Rai D, Nordeng HME. Making fair comparisons in pregnancy medication safety studies: an overview of advanced methods for confounding control. Pharmacoepidemiol Drug Saf 2018; 27:140–47.

54. Bjørn AM, Nørgaard M, Hundborg HH, Nohr EA, Ehrenstein V. Use of prescribed drugs among primiparous women: an 11-year population-based study in Denmark. Clin Epidemiol 2011; 3:149–56.

55. Smolina K, Hanley GE, Mintzes B, Oberlander TF, Morgan S. 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:e0128312.

56. 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; 28:616–24.

57. Van de Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. The GRoLTS-checklist: guidelines for reporting on latent trajectory studies. Struct Equ Model Multidiscip J 2017; 24:451–67.

58. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. New York: Springer-Verlag, 2009.

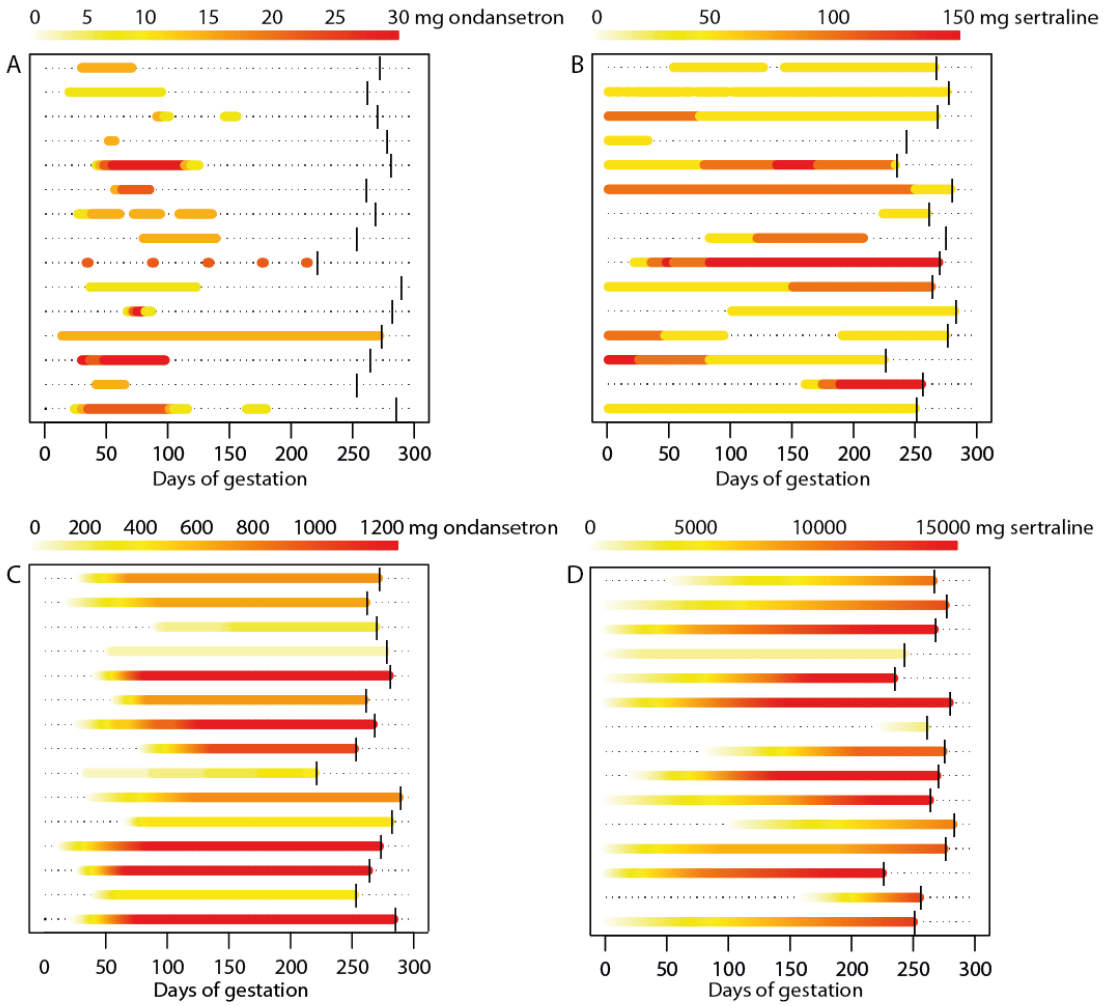


Figure 1. Heat maps of the daily dose (Figures 1A and 1B) and cumulative dose (Figure 1C and 1D) for ondansetron (Figures 1A and 1C) and sertraline (Figures 1B and 1D) use during pregnancy. Each horizontal line represents a woman's pregnancy, with the vertical bar the day of delivery. The dashes in the heat maps of daily dose represent days without use of the medication of interest.

Table 1. Overview of studies on medication use during pregnancy using longitudinal methods for exposure modeling.

Reference	Study population	Exposure of interest	Modeling method	Outcome of interest
Hurault-Delarue *et al*. (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 DDD per month	*K*-means clustering	None
Hurault-Delarue *et al*. (2017) (11)	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 DDD per month	*K*-means clustering	Neonatal pathology: oxygen therapy, intubation, resuscitation, transfer to specialized service, and/or respiratory distress
Bandoli *et al.* (2018) (12)	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 3 months post-partum based on EMR	*K*-means clustering	Birth weight, gestational age at delivery
Palmsten *et al.* (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
Lemon *et al*. (2019) (13)	Liveborn, singleton deliveries at Magee-Womens Hospital with UPMC Health Plan coverage from 2006 through 2014	Ondansetron exposure extracted from the inpatient electronic medical record and through insurance claims for outpatient prescriptions	*K*-means clustering	Neonatal cardiac anomalies
Palmsten *et al*. (2019) (14)	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

Reference	Study population	Exposure of interest	Modeling method	Outcome of interest
Palmsten *et al*. (2020) (15)	Liveborn deliveries between 2012-2016 among females aged 12-49 identified in OptumLabs® Data Warehouse administrative health care claims	Antidepressant pharmacy dispensings from 3 months before LMP through 35 gestational weeks	*K*-means clustering	Preeclampsia and postpartum hemorrhage
Frank *et al*. (2018) (16)	Pregnant women participating in MoBa using hypothyroid medication	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 *et al.* (2019) (17)	Data linked for the Maternal Use of Medications and Safety (MUMS) Study: women who gave birth between 2005-2012 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
Salvatore *et al*. (2017) (10)	Pregnant women participating in MoBa with paracetamol use	Questionnaire: any co-medication used during pregnancy at 4 weeks intervals including indication for use and number of days used	HCA	None
Yonkers *et al*. (2011) (18)	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 *et al*. (2012) (19)	Vaccine and Medication in Pregnancy Surveillance System H1N1 vaccine in pregnancy study: women enrolled before 20 weeks of gestation, USA	H1N1 vaccine	Time-varying approach in Cox proportional hazard models	Miscarriage

Reference	Study population	Exposure of interest	Modeling method	Outcome of interest
Matok *et al*. (2014) (20)	UK HES database linked to the CPRD: women between 15 and 45 years of age who delivered a singleton live birth between April 1, 1997 and March 31, 2012	Decongestant prescriptions between gestational week 27-37 registered in CPRD	Time-varying approach (considered unexposed until prescription) in Cox proportional hazard models	Preterm birth
Daniel *et al*. (2015) (21,22)	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
Bodnar *et al*. (2004) (23)	Iron Supplementation Study: women less than 20 weeks pregnant at the initial visit to a public prenatal clinic in Raleigh, North Carolina, 1997-1999	Randomized 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 *et al*. (2016) (24)	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 3 years of age
Lupattelli *et al*. (2017) (25)	Depressed pregnant women participating in MoBa	Questionnaire antidepressant use during pregnancy, categorized in 4-week intervals	Marginal structural models	Preeclampsia
Lupattelli *et al*. (2018) (26)	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

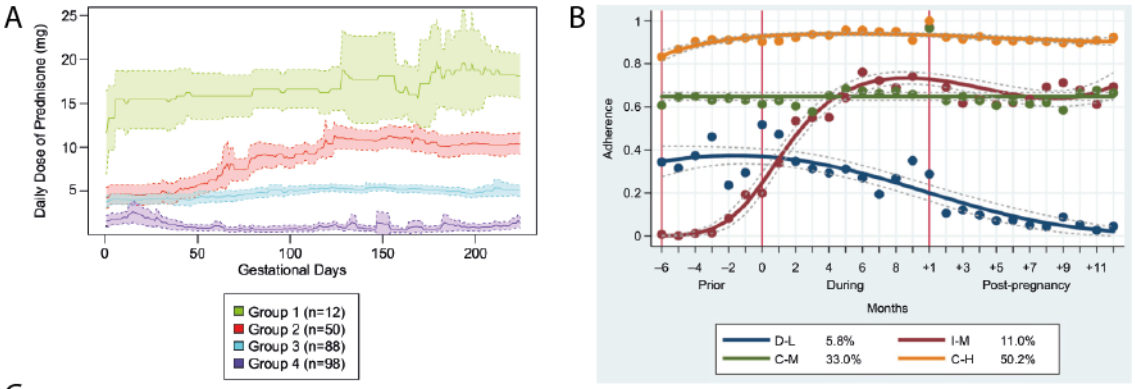
Reference	Study population	Exposure of interest	Modeling method	Outcome of interest
Petersen *et al*. (2018) (27)	Pregnant women participating in the DNBC or MoBa	Paracetamol, aspirin and ibuprofen DNBC: Three telephone interviews, reported on a week-by-week basis MoBa: Questionnaires, reported in 4-week intervals	Marginal structural models	Cerebral palsy

BC, British Columbia; CPRD, Clinical Practice Research Datalink; DDD, Defined Daily Dose; DNBC, Danish National Birth Cohort; EMR, electronic medical records; GBTM, group-based trajectory models; HCA, hierarchical cluster analysis; HES, Hospital Episodes Statistics; LMP, last menstrual period; MoBa, Norwegian Mother and Child Cohort study; NorPD, Norwegian Prescription Database; NSAID, non-steroidal anti-inflammatory drug; SSRI, selective serotonin reuptake inhibitor; UPMC, University of Pittsburgh Medical Center

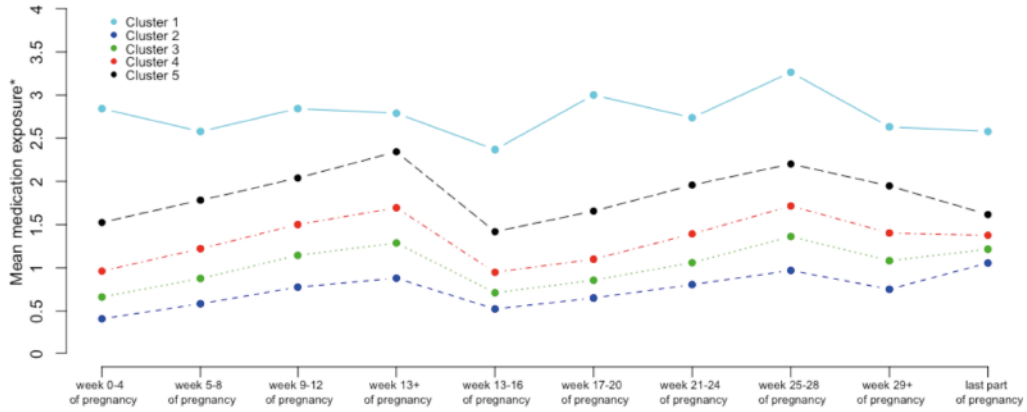
Table 2. Summary of the main applications, advantages, and limitations of *k*-means longitudinal clustering, group-based trajectory modeling, hierarchical cluster analysis, extended Cox models, and g methods

Method	Specified by researcher	Applicability	Advantage	Limitations	
Methods using unsupervised clustering					
K means longitudinal clustering	Number of clusters	Model similar patterns of values for longitudinally collected variable(s)	Non-parametric; requires no assumptions about trajectory shape; optimizes an objective function (minimizing sum of squared error)	Assumptions of equal variances for *k* groups may fail to identify smaller groups; assumes clusters are linearly separable; will identify distinct groups in uniform data	For all unsupervised clustering methods: no guarantee that identified clusters are heuristically or clinically meaningful; vulnerability to biases (confounding, selection, misclassification) may be less apparent
Group-based trajectory models (GBTM)	Number and shape of trajectories; type of parametric model	Finite mixture model for assigning individuals to longitudinal trajectories, given similar values on variable(s) 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 (HCA)	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 datasets	
Methods using a priori exposure definitions					
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	A priori definitions for exposure may not capture the most common patterns or the clinically relevant window of vulnerability

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



C



Clusters	Pregnancies**		Mean medication exposure	Only paracetamol		Use of psychotropics[$]		Use of opioids		Use of NSAIDs[†]		Use of anti-infective or antiseptics		Use of asthma medications (R03)	
	N	%		N	%	N	%			N	%	N	%	N	%
Cluster 1 - "high intensity use"	19	0.2	2.8	0	0.0	10	52.6	7	36.8	11	57.9	17	89.5	7	36.8
Cluster 2 - "low intensity use"	4900	50.1	0.7	187	3.8	357	7.3	186	3.8	1125	22.9	1682	34.3	218	4.4
Cluster 3 - "moderate intensity use"	2643	27.0	1.0	0	0.0	359	13.4	222	8.4	651	24.6	1508	57.1	383	14.5
Cluster 4 - "moderate intensity use, more mental illness"	1811	18.5	1.3	0	0.0	392	21.6	266	14.7	707	39.0	1202	66.4	423	23.4
Cluster 5 - "high intensity use, more asthma"	405	4.1	1.8	0	0.0	145	35.8	107	26.4	200	49.4	313	77.3	152	37.5

*The mean medication exposure is the average number of medications reported for different indications, e.g. if a woman has a mean medication exposure of 2 in week 0-4 of pregnancy, it indicates that in week 0-4 of pregnancy, either she has been exposed to two medications for the same indication or she has been exposed to the same medication twice for treating two different indications

**All mothers included in the study have used paracetamol at each reported time window

[$]Psychotropic medications included anti-depressants (ATC code N06A), anti-epileptics (ATC code N03A), benzodiazepines (ATC codes N05CD and N05BA) and triptans (ATC code N02CC)

[†]Non steroidal anti-inflammatory drugs (NSAIDs), ATC codes M01A, M02A and N02BA

Figure 2. Data visualization of unsupervised clustering methods: *k*-means clustering (A, Palmsten *et al*. (4)), group-based trajectory models (B, Frank *et al*. (16)), and hierarchical cluster analysis (C, Salvatore *et al*. (10)).

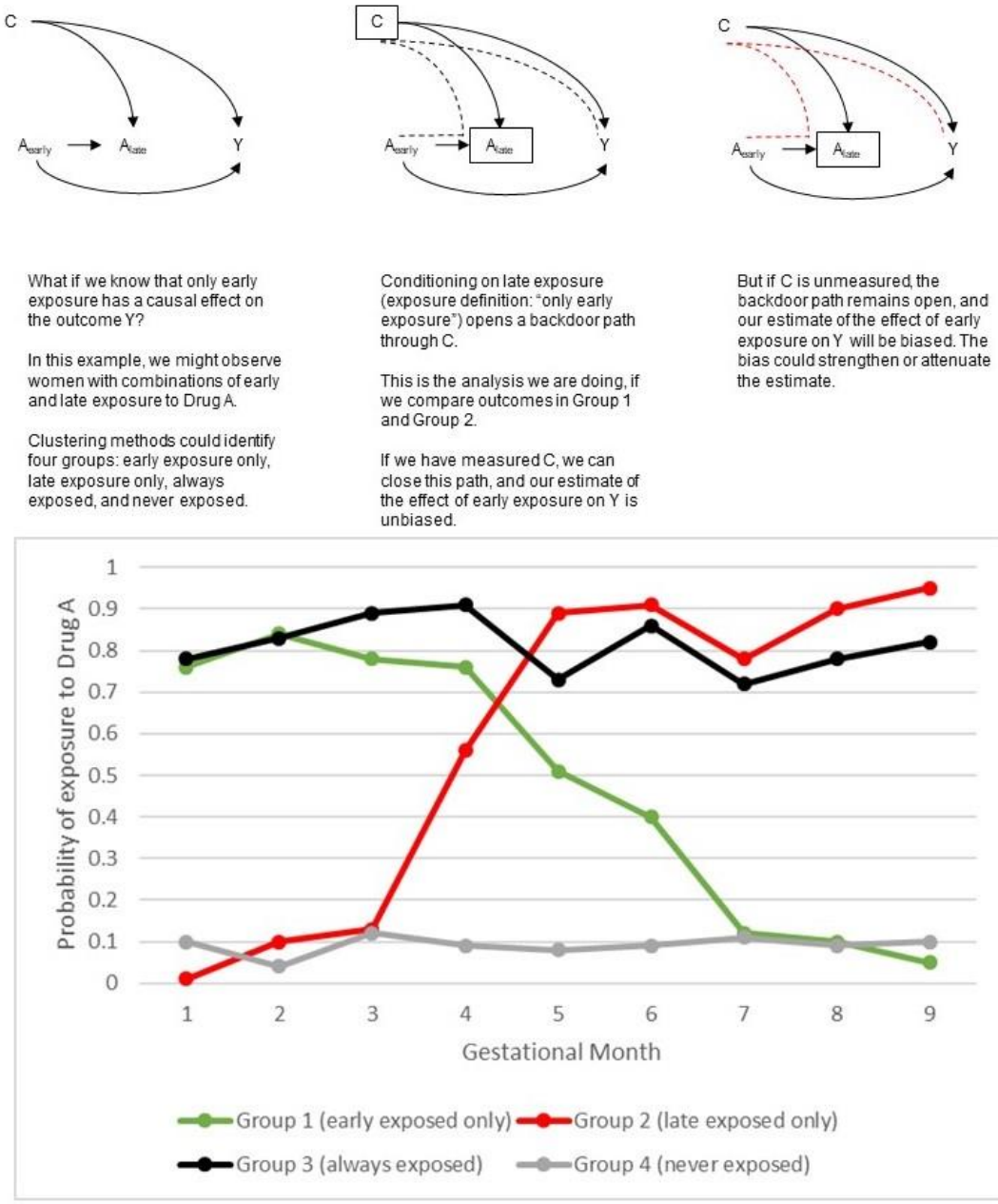


Figure 3. Illustration of potential bias due to conditioning on a collider.

SUPPLEMENTARY MATERIAL

Longitudinal Methods for Modeling Exposures in Pharmacoepidemiologic Studies in Pregnancy

Software included in Supplementary Table 1 is intended as an overview of some currently available programs, rather than a complete list of software solutions.

Supplementary Table 1. Software for methods to model time-varying exposures and longitudinal cluster analyses.

Method	Software
K-means clustering	R: k-means function or kml package; kml3D package for joint trajectories of >1 variable
Group-based trajectory models	Stata plugin *traj* SAS: PROC TRAJ add-on package
Hierarchical cluster analysis	Python: SciPy package R: hclust package SAS: PROC CLUSTER SPSS: hierarchical cluster method Stata (v.13 or later): cluster command
Extended Cox models	Any statistical software package
G methods	SAS macro GFORMULA3.0 Stata: user-written command gformula Possible to implement without specialized software solutions[a]

[a] Many worked examples in SAS, Stata, and R, using publicly available data are published in Hernán MA, Robins JM. Causal inference: what if. Boca Raton: Chapman & Hall/CRC, 2020.