American Journal of Epidemiology

Vol. 190, No. 6
DOI: 10.1093/aje/kwaa288
Advance Access publication:
April 29, 2021

Practice of Epidemiology

Active Surveillance of the Safety of Medications Used During Pregnancy

Krista F. Huybrechts*, Martin Kulldorff, Sonia Hernández-Díaz, Brian T. Bateman, Yanmin Zhu, Helen Mogun, and Shirley V. Wang

* Correspondence to Dr. Krista F. Huybrechts, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street, Boston, MA 02120 (e-mail: khuybrechts@bwh.harvard.edu).

Initially submitted March 4, 2020; accepted for publication December 23, 2020.

The scientific community relies on postmarketing approaches to define the risk of using medications in pregnancy because information available at the time of drug approval is limited. Most studies carried out in pregnancy focus on a single outcome or selected outcomes. However, women must balance the benefit of treatment against all possible adverse effects. We aimed to apply and evaluate a tree-based scan statistic data-mining method (TreeScan; Martin Kulldorff, Harvard Medical School, Boston, Massachusetts) as a safety surveillance approach that allows for simultaneous evaluation of a comprehensive range of adverse pregnancy outcomes, while preserving the overall rate of false-positive alerts. We evaluated TreeScan with a cohort design and adjustment via propensity score techniques, using 2 test cases: 1) opioids and neonatal opioid withdrawal syndrome and 2) valproate and congenital malformations, implemented in pregnancy cohorts nested within the Medicaid Analytic eXtract (January 1, 2000–December 31, 2014) and the IBM MarketScan Research Database (IBM, Armonk, New York) (January 1, 2003–September 30, 2015). In both cases, we identified known safety concerns, with only 1 previously unreported alert at the preset statistical alerting threshold. This evaluation shows the promise of TreeScan-based approaches for systematic drug safety monitoring in pregnancy. A targeted screening approach followed by deeper investigation to refine understanding of potential signals will ensure that pregnant women and their physicians have access to the best available evidence to inform treatment decisions.

drug safety; opioids; pregnancy; systematic surveillance; TreeScan; valproate

Abbreviations: ICD-9, *International Classification of Diseases, Ninth Revision*; MLCC, Multi-Level Clinical Classifications; PS, propensity score; RR, relative risk.

Pregnant women are de facto excluded from most drug-related clinical trials in an effort to protect the fetus from research-related risks (1). It is well known, however, that neither a drug's structure and function nor animal studies are good predictors of human teratogenesis (2, 3). Hence, when a new drug enters the market, there is little or no available information about the safety of its use during pregnancy, and the evidence base regarding safety must rely on postmarketing surveillance.

Historically, spontaneous reporting, pregnancy exposure registries, and case-control study designs were the main approaches used to evaluate the safety of medications in pregnancy. More recently, there has been increasing reliance on cohort studies nested within large health-care utilization databases—i.e., claims databases, national and provincial registries, and electronic health record databases. Such studies offer a number of important advantages, including: 1) the use of large population-based cohorts; 2) prospectively collected exposure information, eliminating the potential for recall bias; 3) availability of internal reference groups; 4) the ability to study a broad range of maternal and infant outcomes; and 5) rich information for confounding adjustment (4). While there are important methodological challenges associated with using this type of data for research, with careful design these data can be a valid source for evaluating the safety of medications in pregnancy, and their use has

been increasingly embraced by the US Food and Drug Administration, the European Medicines Agency, and other regulators.

Across all of these designs, most studies of drug safety in pregnancy focus on a single outcome or selected outcomes, either by design or as a result of selective publication of associations in the context of multiple comparisons. To make an informed treatment decision, a pregnant woman must balance the benefits of treatment against all possible adverse effects for herself and her newborn infant. There is therefore a need to develop a safety surveillance approach that allows for the simultaneous evaluation of a comprehensive range of adverse maternal, fetal, and neonatal outcomes, while accounting for multiple testing and preserving the overall rate of false-positive alerts.

Our objective in this study was to develop such an approach by building on the tree-based scan statistic method (TreeScan; Martin Kulldorff, Harvard Medical School, Boston, Massachusetts) (5), which has been used by the Food and Drug Administration and the Centers for Disease Control and Prevention to study the safety of drugs and vaccines outside of pregnancy (6–8). We evaluated the approach using 2 test cases of medications known to cause specific adverse pregnancy outcomes: 1) prescription opioids and the risk of neonatal opioid withdrawal syndrome and 2) valproate and the risk of specific congenital malformations.

METHODS

The tree-based scan statistics for drug safety surveillance scan a hierarchical tree of specific outcomes, as well as groups of clinically related outcomes, for associations with the exposure of interest (6, 9). It does so while accounting for the multiple testing of correlated hypotheses. As such, the method allows for the simultaneous evaluation of a broad range of potential adverse events, as well as groupings of related adverse events. Because an adverse event "signal" does not necessarily mean that there is a causal relationship, such an approach represents the first step in a comprehensive drug safety surveillance system, used to highlight potential problems that warrant further, thorough investigation. Below, we describe a TreeScan approach for drug safety surveillance during pregnancy and then detail the test cases used to evaluate the feasibility of the approach.

TreeScan

The "tree" in TreeScan refers to a classification system that hierarchically groups coded clinical concepts into clinically related categories. Example clinical coding systems that could be used include, among others, the *International Classification of Diseases*, the Multi-Level Clinical Classifications (MLCC) (10) for *International Classification of Diseases* codes, or the Medical Dictionary for Regulatory Activities classification system (MedDRA MSSO, McLean, Virginia), where each grouping of clinically related concepts represents an outcome "node" in the hierarchical tree. In pregnancy research, tree structures can be customized to focus on outcomes for which exposure may be in an etiologically relevant window. For example, tree structures that involve congenital malformation outcomes may be relevant when studying exposures incurred in the first trimester but not the third trimester. A tree that allows a broader scan of potential maternal and neonatal outcomes other than malformations may be more appropriate for exposures occurring later in pregnancy. Examples of the hierarchical tree structures used in our test cases are shown in Figure 1 and Figure 2 (see "Test cases" below for details). Use of a hierarchical tree structure maximizes statistical power to detect clinically related outcomes, which would not be feasible by scanning only individual uncommon diagnoses.

The "scan" in TreeScan refers to the statistic used to evaluate potential associations between the exposure and the range of outcomes in the hierarchical tree. Tree-based scan statistics are available with different probability models for different types of data, including the conditional and unconditional versions for Bernoulli/binomial and Poisson-generated data (5–7, 9, 11). These statistics test the null hypothesis of no difference in risk of adverse events in any outcome node in the tree against a 1-sided alternative that there is at least 1 node in the tree where the risk of adverse events is higher in the exposed group than in the comparator group.

When screening for multiple outcomes, it is important to control the rate of false-positive alerts. Therefore, TreeScan generates multiplicity-adjusted P values that accurately reflect the type I error rate in the absence of confounding. P values can be interpreted at face value as the probability of seeing an association of the observed magnitude or one more extreme if the null hypothesis were true. The distribution of the tree-based scan statistic is unknown and is therefore derived nonparametrically by generating distributions under the null hypothesis of no effect of exposure in any node via Monte Carlo simulation. The test statistics from 9,999 data sets simulated under the null and from the 1 observed data set are ranked from largest to smallest. The multiple-testing–adjusted P value is determined by the rank R of the observed test statistic divided by 10,000 (9,999 simulated + 1 observed data set), so that $P = R/(9{,}999 + 1)$.

Propensity score (PS) matching can be used for confounding adjustment with a Bernoulli tree-based scan statistic. However, this test statistic is not compatible with weighting-based adjustment methods. In contrast, fine-stratification weights can be used to handle confounding with the Poisson tree-based scan statistic. Fine stratification on the PS is a strategy often used in pregnancy studies that focus on evaluation of the risk of a specific outcome (12–14). This adjustment strategy can boost power to detect small associations compared with PS matching and may be particularly useful in settings of rare exposure and outcomes (15). For our data, we used the unconditional Bernoulli and Poisson test statistics.

When screening to prioritize associations that warrant further investigation, one can specify a priori a threshold for statistical alerting, which corresponds to the probability of a type I error (falsely rejecting the null hypothesis that there is no excess risk for any outcome grouping in the tree). In our test cases, we set the threshold for alerting at 5%, a commonly used value in biomedical research. Therefore, only

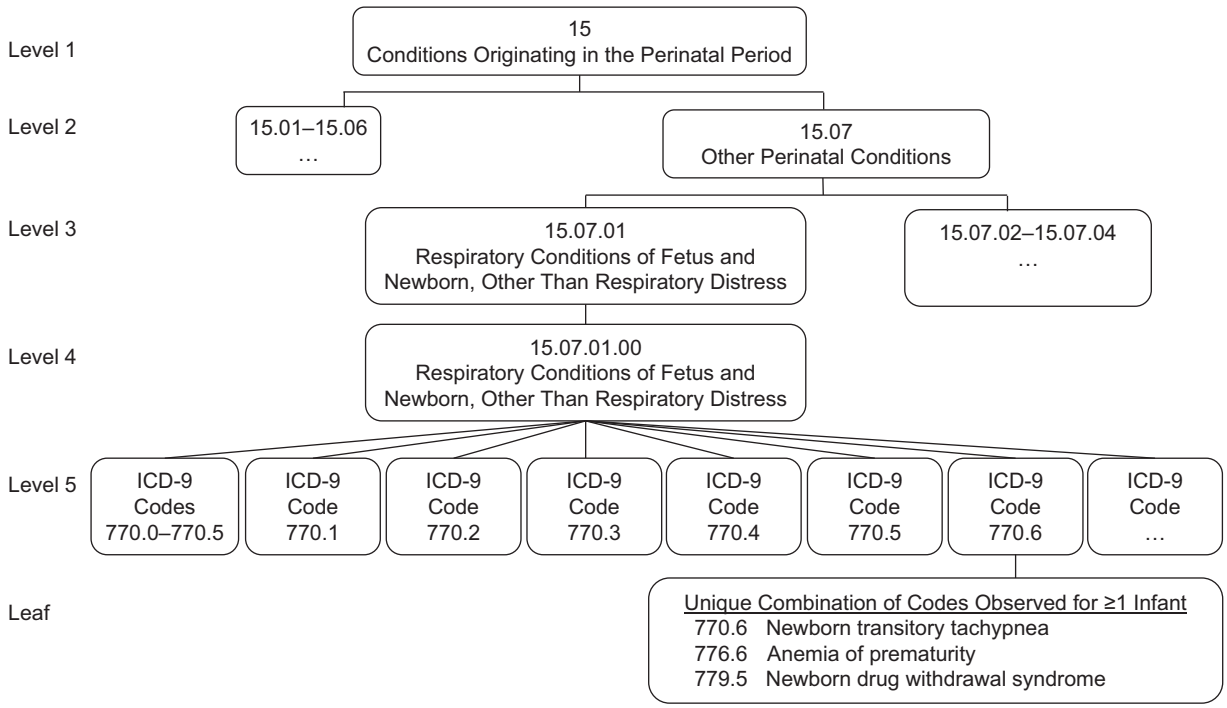

Figure 1. Example of a tree structure for 1 branch of the Multi-Level Clinical Classifications (MLCC) software tree. The MLCC tree is organized by body system, and condition categories are defined by the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project. The leaf node is comprised of unique combinations of diagnostic codes observed in the follow-up window. Each code within the leaf has a unique level 5 node, the individual *International Classification of Diseases, Ninth Revision* (ICD-9) code. This level 5 code is grouped at increasingly aggregated higher levels by condition and/or body system. For example, if an infant had 3 diagnoses within 90 days of delivery (ICD-9 codes 770.6 (newborn transitory tachypnea), 776.6 (anemia of prematurity), and 779.5 (newborn drug withdrawal syndrome)), this unique combination of codes would form the leaf level of the tree. Each diagnosis code in that unique combination would be grouped in clinically related categories based on the MLCC categories. ICD-9 code 770.6 would be grouped with other codes related to respiratory conditions of the fetus and newborn at levels 4 and 3. It would then be further grouped with other level 3 nodes to create a level 2 node encompassing "other perinatal conditions." This level 2 node is further grouped with other level 2 nodes to form the level 1 node of "conditions originating in the perinatal period." A similar grouping process would occur for the other ICD-9 codes in the leaf-level combination of codes assigned to the infant.

outcome nodes with $P < 0.05$ were considered statistical alerts. This means that if the null hypothesis is true, there is a 95% probability of having 0 alerts and a 5% probability of having 1 or more false alerts.

Adverse pregnancy outcomes tend to be rare. Because multiple-testing adjustment reduces power to detect true effects, some statistical nonalerts with high relative risk may still merit assessment using a clinical and methodological perspective to determine whether these outcomes warrant further consideration. Thus, these screening methods can play an important part in prioritization of potential signals even when there is not sufficient power to alert at a stringent prespecified statistical threshold.

Relatedly, a statistical alert does not necessarily equal a safety signal. Statistical alerts help prioritize associations that are unlikely to have occurred by chance. Residual confounding not accounted for in the design phase can produce spurious alerts. Potential signals of concern should therefore be followed by a cohort study tailored to the specific drug and outcome of interest, first using the original data source to assess whether the observed association remains or is removed with a tailored design and confounding adjustment. For associations that persist, the robustness of the finding should be further evaluated by implementing the study in independent data.

All analyses were conducted using publicly available TreeScan software (5).

Test cases

Data sources and study cohorts. We used linked mother-infant pregnancy cohort data from the Medicaid Analytic eXtract (2000–2014; $n = 1{,}991{,}722$) (16) and the IBM MarketScan Research Database (IBM, Armonk, New York) (2003–2015; $n = 1{,}324{,}803$). The pregnancy cohorts consisted of all pregnancies among women aged 12–55 years resulting in live birth with continuous insurance coverage from 3 months before the start of pregnancy to a minimum of 1 month after delivery. We required offspring to have continuous insurance coverage for a minimum of 3 months after the date of birth, unless they died sooner.

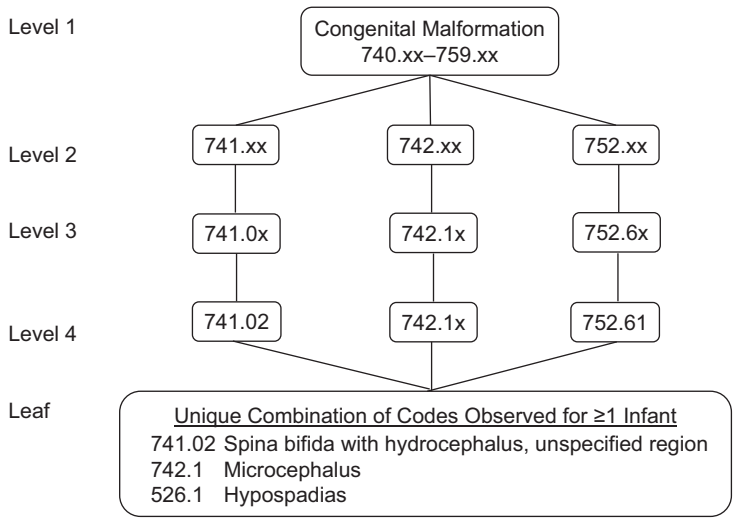

Figure 2. Example of an *International Classification of Diseases, Ninth Revision* (ICD-9) digit-based tree structure for 1 patient's unique combination of congenital malformations. The leaf node is comprised of unique combinations of diagnostic codes observed in the follow-up window. Each code within the leaf has a unique level 4 node, a 5-digit ICD-9 code (with a placeholder if there is no fifth digit code available). This level 4 code is grouped at increasingly aggregated higher levels based on 4-digit codes, 3-digit codes, and then, finally, any congenital malformation code at the highest level of the hierarchical tree.

The exposures for the 2 example studies of interest were selected for their reasonably well-characterized safety profiles. Our goal was to evaluate whether the known risks associated with the medication would be detected using the TreeScan approach, without generating many false-positive alerts (Table 1).

Test case 1: opioid exposure in late pregnancy. Our objective in the first test case was to evaluate the risk of neonatal outcomes associated with late-pregnancy exposure to prescription opioids. Late-pregnancy opioid exposure is known to be associated with neonatal opioid withdrawal syndrome (17, 18), and we therefore anticipated detecting a signal for this outcome. Less is known about the potential risk of other adverse neonatal outcomes. This example is characterized by a relatively common exposure and a relatively common known safety concern.

Women were considered exposed if they filled at least 2 prescriptions for opioids during the 60 days before delivery (see Web Figure 1 and Web Table 1, available online at https://doi.org/10.1093/aje/kwaa288). Women were considered unexposed if they did not fill any prescriptions for opioids, buprenorphine, or methadone from 3 months before the start of pregnancy to the end of pregnancy. We considered a broad range of potential confounding variables, including maternal demographic characteristics, opioid indications, substance use disorder or dependence, obstetrical conditions, chronic comorbid conditions, health-care utilization, and concomitant medication use (Web Table 2). Covariates were measured from 180 days before delivery to the end of pregnancy. A PS was estimated on the basis of a logistic regression model including all prespecified covariates.

We used a 1:5 PS matching approach (nearest-neighbor algorithm with a matching caliper of 0.01) to adjust for confounding and used the unconditional Bernoulli tree-based scan statistic to evaluate whether there were outcomes or groupings of outcomes that were more likely to occur after late-pregnancy exposure to opioids.

We used an MLCC-based (10) tree to scan across neonatal outcomes. The MLCC system groups *International Classification of Diseases, Ninth Revision* (ICD-9) diagnosis codes (leaf level) into hierarchical levels based on body systems, where each grouping is an outcome "node" in the tree. We conducted hypothesis testing for 9,044 hierarchical outcome nodes at every level of the tree above the leaf level, without double-counting outcomes when infants had multiple ICD-9 diagnoses grouped in the same higher-level aggregated outcome nodes of the hierarchical tree (see Figure 1 for an example). The follow-up window for identifying infant outcomes was from birth through age 90 days. Prior TreeScan work conducted outside of pregnancy (6–8, 11, 19, 20) has used MLCC-based trees to scan across thousands of correlated outcomes.

Pruning the tree to avoid hypothesis-testing on irrelevant nodes can help increase power to detect associations between exposure and the remaining outcomes. Because we evaluated risks associated with late-pregnancy exposure to opioids, which falls after the etiologically relevant exposure risk window for most congenital malformations, branches with codes related to congenital malformations were removed. The tree was also "pruned" to remove diagnosis codes unlikely to reflect an adverse reaction caused by drugs (e.g., a well care visit or live birth), diagnoses that did not represent incident events (e.g., a family history of alcoholism), and conditions with long latency/induction periods (e.g., cancer). The pruned tree is available from the authors upon request.

For the opioids, the unconditional Bernoulli TreeScan model was used.

Table 1. Comparison of the Specific TreeScan[a] Methods Used for Safety Surveillance in 2 Test Cases, Implemented in Pregnancy Cohorts Nested in the Medicaid Analytic eXtract and the IBM MarketScan Research Database[b]

Parameter	Test Case 1: Opioids	Test Case 2: Valproate
Exposure	Late pregnancy exposure, relatively common	First-trimester exposure, rare
Expected outcome	Neonatal opioid withdrawal syndrome, relatively common	Specific birth defects, rare
Confounding adjustment method	Propensity score matching	Propensity score fine stratification
Scan statistic	Unconditional Bernoulli	Unconditional Poisson
Hierarchical outcome tree	Pruned Multi-Level Clinical Classifications Software, no birth defects	ICD-9 digit-based, only birth defects
Washout to identify incident outcomes	0 days	0 days
Outcome counts	Any unique occurrence of a code in any care setting or diagnosis position within 90 days on or following delivery	Any unique occurrence of a code in any care setting or diagnosis position within 90 days on or following delivery

Abbreviation: ICD-9, International Classification of Diseases, Ninth Revision.
[a] Martin Kulldorff, Harvard Medical School, Boston, Massachusetts.
[b] IBM, Armonk, New York.

Test case 2: valproate exposure in the first trimester of pregnancy. The objective of the second test case was to evaluate the risk of congenital malformations associated with first-trimester exposure to valproate. The teratogenicity of valproate has been well documented (21). Aside from congenital malformations overall, valproate exposure has been associated with malformations such as neural tube defects, cardiac defects, orofacial/craniofacial defects (e.g., cleft palate), hypospadias, and skeletal and limb malformations. A clear dose-response relationship has been established (21). This example is characterized by an uncommon exposure and uncommon known safety concerns.

We excluded pregnancies with a chromosomal abnormality, as well as pregnancies with exposure to known teratogens (i.e., warfarin, antineoplastic agents, lithium, isotretinoin, misoprostol, and thalidomide) (Web Figure 2). Women were considered exposed if they filled a prescription for valproate during the first trimester of pregnancy (i.e., divalproex, valproic acid, or valproate sodium). Women without any filled prescription for valproate or other anticonvulsant medications from 3 months before the start of pregnancy to the end of the first trimester were considered unexposed (n = 3,544 exposed; n = 1,940,120 unexposed). Potential confounders considered included maternal demographic characteristics, indications for anticonvulsant drug use, obesity, smoking, chronic comorbidity (e.g., diabetes mellitus, hypertension), concomitant medications (e.g., antipsychotic agents, antidepressants), and proxies for overall health status (Web Table 3) (22). Covariates were measured from 3 months before the start of pregnancy to the end of the first trimester, except for health-care utilization variables, which were measured during the 3 months before pregnancy only. A PS was estimated on the basis of a logistic regression model including all prespecified covariates.

In order to preserve power, we used a PS fine-stratification (15) approach to adjust for confounding. The unconditional Poisson tree-based scan statistic was used to evaluate whether and which congenital malformations were associated with valproate exposure. After excluding pregnancies in the nonoverlapping areas of the PS distribution, 50 strata were defined on the basis of percentiles of the PS calculated in the valproate-exposed pregnancies. All exposed pregnancies received a weight of 1, and all unexposed pregnancies were weighted in proportion to the distribution of the exposed in the stratum into which they fell (15).

Because the MLCC categories for congenital malformations are very coarse, we developed a tree based on ICD-9 codes focused only on congenital malformation codes from mother or infant claims. The top level was defined by the occurrence of a congenital malformation code in any care setting and diagnosis position (ICD-9 codes 740.x–759.x). The next 3 levels were increasingly specific and defined by 3-digit, 4-digit, and 5-digit ICD-9 codes. The final, "leaf" level was defined by the unique combination of ICD-9 congenital malformation codes observed for a mother-baby pair to allow hypothesis-testing at more aggregated levels of the tree without double counting of outcomes (see Figure 2 for an example). We evaluated 708 outcome nodes above the leaf level in the scan.

For valproate, the unconditional Poisson TreeScan model was used.

This research was approved by the institutional review board of Brigham and Women's Hospital, which granted a waiver of informed consent.

RESULTS

Test case 1: opioid exposure in late pregnancy

The source cohort included 53,771 women who filled at least 2 prescriptions for opioids during the 60 days before delivery and 1,360,039 unexposed women. After 1:5 matching on the PS, 24,080 exposed women and 120,400 unexposed women remained in the analytical cohort.

Table 2 lists all nodes with $P < 0.5$ to allow for the identification of nonalerts with high relative risk that may still merit further assessment (Web Table 4 lists all nodes with $P < 1$). The only tree branch on which there were statistical alerts at $P < 0.05$ was the branch related to the expected safety concerns of drug withdrawal in the newborn (Table 2, Web Table 4). Strong alerts emerged for "drug withdrawal syndrome in the newborn" (relative risk (RR) = 6.1), "drug withdrawal" (RR = 4.5), and "narcotics affecting fetus or newborn via placenta or breast milk" (RR = 5.4). There was also a strong alert on this branch for "opioid-type dependence, unspecified use," with a relative risk of 20.0, which is a marker for in-utero exposure to opioids. There were no false-positive alerts at the statistical alerting threshold of 0.05.

Test case 2: valproate exposure in the first trimester of pregnancy

The cohort for the second test case consisted of 3,544 pregnancies with exposure to valproate during the first trimester. The 1,890,250 unexposed pregnancies that remained after removing 49,870 pregnancies in nonoverlapping areas of the PS distribution were used to estimate the expected counts.

Alerts with prior supporting evidence

There were statistical alerts with P values less than 0.05 for several of the known associations (Table 3, Web Table 5). The relative risks for spina bifida (the most common neural tube defect) ranged from 4.2 to 165.1, depending on the level of the outcome node being tested (specific diagnostic code vs. higher-level grouping). Alerts for an increased risk were observed for several cardiac malformation codes (bulbus cordis anomalies and anomalies of cardiac septal closure (RR = 1.4), as well as other congenital anomalies of the circulatory system (RR = 1.4), which was driven by patent ductus arteriosus (RR = 1.6)), for hypospadias (RR = 2.3) and for polydactyly (RR = 2.2). Multiple congenital anomalies (RR = 4.5) and other and unspecified congenital anomalies (RR = 2.3) also alerted at the preset threshold.

Although a 2- to 4-fold increased risk was observed for several oral cleft-related codes, the alert threshold was not reached (Web Table 5).

Unanticipated alerts

Finally, we found a significant alert for indeterminate sex and pseudohermaphroditism (RR = 6.8, $P = 0.02$), which has not previously been reported. However, some valproate formulations contain phthalates as inactive ingredients for gastroprotection. Phthalates have estrogenic effects and have previously been associated with genital malformations (23, 24).

DISCUSSION

Using 2 test cases of medications known to be associated with adverse pregnancy outcomes—prescription opioids and valproate—we evaluated an approach for drug safety surveillance in pregnancy that allows for the simultaneous consideration of a comprehensive range of adverse outcomes, with adjustment for confounding, while preserving the overall rate of false-positive alerts. The approach leverages the established TreeScan statistic approach and extensive prior experience in the perinatal pharmacoepidemiology community with the conduct of drug safety studies in pregnancy nested within large health-care utilization databases. The test cases represented different scenarios in terms of prevalence of the exposure and adverse outcomes of interest, and they allowed us to test positive control associations (i.e., known effects) and negative control associations (i.e., the unknown effects on other outcomes). In both instances, the known safety concerns were identified without generating many additional positive alerts. While the test cases parallel the anticipated areas of interest for signal identification activities in pregnant women, broader implementation of this approach would involve evaluating different exposure windows and all maternal and neonatal outcomes for a given drug of interest.

On the basis of this initial evaluation, TreeScan-based approaches to systematic drug safety monitoring in pregnancy appear promising. As the approach gets implemented more broadly, further refinement of the methods to accommodate the intricacies of drug safety research in pregnancy will be helpful. Possible modifications include—among others—development of outcome trees with hierarchical groupings that optimize the power to detect safety signals (e.g., by grouping outcomes that are anatomically different but pathogenetically similar), improved methods for confounding control when simultaneously evaluating a broad range of outcomes (e.g., high-dimensional PSs), and methods that avoid defining the exposure window in a way that creates differential opportunity for capture of exposure in pregnancies of shorter versus longer duration.

A frequently raised concern when screening for associations is that it is a "hypothesis-free" approach that could result in false-positive signals that would unnecessarily alarm pregnant women and their physicians. However, this concern is misplaced, for 2 reasons. First, it is important to note that the underlying pathophysiology of many adverse pregnancy outcomes is not fully understood, and the biological mechanisms for many accepted human teratogens (including thalidomide) remain unknown. Potential adverse effects identified using a robust screening approach can therefore not be dismissed simply because a known biological explanation has not been established. Second, the TreeScan approach controls the overall error rate. Therefore, if the null hypothesis were true, out of 100 studies that scanned an outcome tree, on average only 1 study would produce a false-positive signal when the type I error α for alerting was set to 0.01, while the remaining 99 studies would have 0 alerts. This is in contrast to current practice, where multiple associations are frequently tested in a single study without adjustment for multiple testing, resulting in a much higher type I error rate than the experimentwide α level. It should also be noted that P values are used as a means to rank and prioritize alerts for further investigation, rather than in the way they are sometimes misused, to decide whether there is a causal association.

Table 2. Safety Alerts for Opioid Exposure in Late Pregnancy, Based on Pregnancy Cohorts Nested Within the Medicaid Analytic eXtract (2000–2014) and the IBM MarketScan Research Database (IBM, Armonk, New York) (2003–2015)[a]

MLCC Node Identifier[b]	Node Description	*P* Value	No. of Outcomes	Risk (*R*)		RR	RD	Observed/ Expected	Observed − Expected
				R_{exp}	R_{ref}				
05	Mental illness	0.001	2,857	72.0	32.6	2.2	39.4	1.5	303.6
05.12	Substance-related disorders	0.001	1,390	52.8	12.3	4.3	40.4	2.3	363.0
05.12.00	Substance-related disorders	0.001	1,390	52.8	12.3	4.3	40.4	2.3	363.0
05.12.00.00	Substance-related disorders	0.001	1,390	52.8	12.3	4.3	40.4	2.3	363.0
779.5	Drug withdrawal syndrome in newborn	0.001	778	35.2	5.8	6.1	29.5	2.8	272.4
760.72	Narcotics affecting fetus or newborn via placenta or breast milk	0.001	458	19.6	3.6	5.4	16	2.6	146.4
292.0	Drug withdrawal	0.001	123	4.8	1.1	4.5	3.7	2.4	33.4
304.00	Opioid-type dependence unspecified use	0.001	15	1.0	0.0	20.0	0.9	4.0	9.0
06	Diseases of the nervous system and sense organs								
06.08	Ear conditions								
06.08.03	Other ear and sense organ disorders								
06.08.03.00	Other ear and sense organ disorders								
389.9	Unspecified hearing loss	0.07	1,002	20.8	12.3	1.7	8.5	1.3	52.6
15	Conditions originating in the perinatal period								
15.07	Other perinatal conditions								
15.07.04	Other and unspecified perinatal conditions								
15.07.04.00	Other and unspecified perinatal conditions								
760.79	Other noxious influences affecting fetus or newborn via placenta or breast	0.15	726	15.5	8.9	1.7	6.6	1.3	42.8

Abbreviations: exp, exposed; MLCC, Multi-Level Clinical Classifications; RD, risk difference; ref, reference; RR, relative risk.

[a] All nodes with $P < 0.5$ are included.

[b] To show the tree hierarchy, aggregated nodes appear in the table when a more specific descendant was ranked, even if the more aggregated node was not ranked.

Table 3. Safety Alerts for Valproate Exposure During the First Trimester of Pregnancy, Based on Pregnancy Cohorts Nested Within the Medicaid Analytic eXtract (2000–2014) and the IBM MarketScan Research Database (IBM, Armonk, New York) (2003–2015)[a]

ICD-9 Node Identifier[b]	Node Description	Recognized Safety Issue[c]	*P* Value	No. of Cases Observed[d]	No. of Cases Expected	RR
740.x–759.x	Any congenital malformation	Any malformation	0.005	631	528.7	1.2
741	Spina bifida	Neural tube defect	0.14	11	3.4	3.3
741.0	Spina bifida with hydrocephalus	Neural tube defect	0.002	<11	1.2	7.5
741.00	Spina bifida unspecified region with hydrocephalus	Neural tube defect	0.005	<11	1.1	7.5
741.02	Spina bifida dorsal (thoracic) region with hydrocephalus	Neural tube defect	0.01	<11	0.0	165.1
741.03	Spina bifida lumbar region with hydrocephalus	Neural tube defect	0.005	<11	0.5	12.8
741.9	Spina bifida without mention of hydrocephalus	Neural tube defect	0.02	11	2.6	4.2
741.90	Spina bifida unspecified region without hydrocephalus	Neural tube defect	0.02	<11	2.2	4.6
741.93	Spina bifida lumbar region without hydrocephalus	Neural tube defect	0.05	<11	0.9	6.7
745	Bulbus cordis anomalies and anomalies of cardiac septal closure	Cardiovascular defect	0.02	181	133.0	1.4
745.5	Ostium secundum type atrial septal defect	Cardiovascular defect	0.17	150	113.6	1.3
747	Other congenital anomalies of circulatory system	Cardiovascular defect	0.02	144	102.9	1.4
747.0	Patent ductus arteriosus	Cardiovascular defect	0.002	107	65.6	1.6
752	Congenital anomalies of genital organs	Hypospadias	0.27	77	52.9	1.5
752.61	Hypospadias	Hypospadias	0.005	36	15.6	2.3
752.7	Indeterminate sex and pseudohermaphroditism		0.02	<11	1.0	6.8
755.0	Polydactyly	Limb and skeletal defects	0.04	28	12.9	2.2
755.00	Polydactyly unspecified digits	Limb and skeletal defects	0.13	11	3.3	3.3
755.01	Polydactyly of fingers	Limb and skeletal defects	0.06	24	10.6	2.3
755.29	Longitudinal deficiency phalanges, complete or partial, finger	Limb and skeletal defects	0.47	<11	0.3	9.6
759	Other and unspecified congenital anomalies		0.007	33	14.1	2.3
759.7	Multiple congenital anomalies so described		0.001	16	3.5	4.5

Abbreviations: ICD-9, *International Classification of Diseases, Ninth Revision*; RR, relative risk.
[a] All nodes with $P < 0.5$ are included.
[b] To show the tree hierarchy, aggregated nodes appear in the table when a more specific descendant was ranked, even if the more aggregated node was not ranked.
[c] All safety alerts relate to recognized safety issues.
[d] Cell sizes less than 11 for observed outcomes are suppressed in accordance with the Centers for Medicare and Medicaid Services cell size suppression policy.

The cost of controlling the type I error rate is a loss of power to detect true safety signals, raising concerns about false-negative signals. The multiplicity adjustment for TreeScan, however, is less conservative than that for other methods that do not account for the correlation between hypotheses being tested (e.g., Bonferroni correction). Second, concerns about false-negative signals highlight the importance of developing outcome trees specifically targeted toward pregnancy outcomes (as opposed to the more generic MLCCS or ICD-9–based trees), with hierarchical groupings informed by embryology or shared disease processes in order to maximize the power to detect safety signals. Depending on the exposure window considered (e.g., periconception, first trimester, late pregnancy), the tree can be "pruned" by removing outcomes that are known to not be meaningful for a given exposure window, and by removing diagnosis codes unlikely to be caused by drug exposure during pregnancy (e.g., codes related to well care visits or family history). Finally, there is no need to strictly focus on a statistical significance threshold for alerting. Outcomes that do not alert at the prespecified threshold may still have a relatively low likelihood under the null hypothesis (i.e., higher-than-expected incidence). Thus, the method can play an important role in screening and prioritization even if there is not sufficient power to alert at a stringent prespecified threshold, by painting a clinical picture of the pattern of outcomes that are unlikely to be observed if there was no relationship with exposure.

While TreeScan ensures that the rate of false-positive alerts due to chance is controlled as strictly as in a regular epidemiologic study, false alerts due to confounding remain a possibility. When simultaneously evaluating multiple potential adverse events, it is impossible to adjust for confounding as thoroughly as can be done in a study evaluating a single exposure-outcome relationship. In addition, there is a higher possibility of outcome misclassification when scanning across diagnostic codes, compared with a study that uses a highly specific, preferably validated, outcome definition. Therefore, alerts for outcomes and outcome clusters of potential interest should be identified and further evaluated in follow-up studies with confounding adjustment tailored to the exposure and outcome under investigation and using outcome definitions based on validated algorithms. Decisions about alerts worthy of further exploration should be based not only on *P* values (which conflate sample size and effect size) but also on the observed relative and attributable risks, disease severity, and other clinical criteria.

The systematic monitoring of pregnancy-related drug adverse events is important, not only to quickly detect problems when they exist but also to show the absence of strong harmful effects across a range of potential outcomes when there are none, so that pregnant women do not refrain from taking helpful drugs due to a lack of knowledge about their safety profile. Broad implementation of a targeted screening approach like TreeScan will ensure that pregnant women and their physicians have timely access to the best available evidence on the safety of medications in pregnancy, allowing them to weigh the benefit of treatment against all possible risks for both themselves and their offspring. It will be important for the evidence to be updated as data accumulate over time to increase our confidence in the observed associations and reduce the likelihood of false-negative signals.

ACKNOWLEDGMENTS

Author affiliations: Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States (Krista F. Huybrechts, Martin Kulldorff, Brian T. Bateman, Yanmin Zhu, Helen Mogun, Shirley V. Wang); Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States (Krista F. Huybrechts, Sonia Hernández-Díaz); and Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States (Brian T. Bateman).

This study was supported by internal funds from the Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School.

K.F.H. has been an investigator on research grants awarded to Brigham and Women's Hospital from Eli Lilly and Company (Indianapolis, Indiana) and Takeda Pharmaceutical Company (Boston, Massachusetts) for unrelated studies. S.H.-D. has been an investigator on research grants awarded to her institution from Takeda for unrelated studies; has received personal fees from UCB S.A. (Brussels, Belgium) outside the scope of this work; and has served as an epidemiologist with the North American AED Pregnancy Registry (Boston, Massachusetts), which is funded by multiple companies. B.T.B. has been an investigator on research grants awarded to Brigham and Women's Hospital from Baxalta, Inc. (Bannockburn, Illinois), Pacira BioSciences, Inc. (Parsippany, New Jersey), and Takeda for unrelated studies; has received personal fees from Aetion, Inc. (New York, New York) and the Alosa Foundation (Boston, Massachusetts) outside the scope of this work; and has served on an expert panel for a postpartum-hemorrhage quality improvement project that was conducted by the Association of Women's Health, Obstetric, and Neonatal Nurses and funded by a grant from Merck for Mothers (Merck & Co., Inc. (Kenilworth, New Jersey)). S.V.W. received salary support from grants awarded to Brigham and Women's Hospital from Boehringer Ingelheim (Ingelheim am Rhein, Germany), Novartis Pharmaceuticals Corporation (East Hanover, New Jersey), and Johnson & Johnson (Titusville, New Jersey) for unrelated work. The other authors report no potential conflicts of interest.

REFERENCES

1. Wisner KL. The last therapeutic orphan: the pregnant woman. *Am J Psychiatry*. 2012;169(6):554–556.

2. Wilson JG. Evaluation of human teratologic risk in animals. In: Lee DH, Hewson EW, Okun D, eds. *Environment and Birth Defects*. 1st ed. New York, NY: Academic Press, Inc.; 1973:146–160.
3. Brent RL. Utilization of animal studies to determine the effects and human risks of environmental toxicants (drugs, chemicals, and physical agents). *Pediatrics*. 2004; 113(4 suppl):984–995.
4. Huybrechts K, Bateman B, 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.
5. Kulldorff M. TreeScan: software for the tree-based scan statistic. https://www.treescan.org/. Published 2014. Accessed March 4, 2020.
6. Kulldorff M, Dashevsky I, Avery TR, et al. Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol Drug Saf*. 2013;22(5):517–523.
7. Wang SV, Maro JC, Baro E, et al. Data mining for adverse drug events with a propensity score matched tree-based scan statistic. *Epidemiology*. 2018;29(6):895–903.
8. Yih WK, Maro JC, Nguyen M, et al. Assessment of quadrivalent human papillomavirus vaccine safety using the self-controlled tree-temporal scan statistic signal-detection method in the Sentinel System. *Am J Epidemiol*. 2018; 187(6):1269–1276.
9. Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. *Biometrics*. 2003;59(2): 323–331.
10. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Clinical Classifications Software for Services and Procedures. User guide and other resources. https://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp. Accessed April 3, 2020.
11. Maro JC, Nguyen MD, Dashevsky I, et al. Statistical power for postlicensure medical product safety data mining. *eGEMS*. 2018;5(1):Article 6.
12. Huybrechts KF, Hernández-Díaz S, Straub L, et al. Association of maternal first-trimester ondansetron use with cardiac malformations and oral clefts in offspring. *JAMA*. 2018;320(23):2429–2437.
13. Zhu Y, Bateman BT, Gray KJ, et al. Oral fluconazole use in the first trimester and risk of congenital malformations: population based cohort study. *BMJ*. 2020;369:m1494.
14. Huybrechts KF, Bateman BT, Pawar A, et al. Maternal and fetal outcomes following exposure to duloxetine in pregnancy: cohort study. *BMJ*. 2020;368:m237.
15. Desai RJ, Rothman KJ, Bateman BT, et al. Propensity-score-based fine stratification approach for confounding adjustment when exposure is infrequent. *Epidemiology*. 2017;28(2):249–257.
16. Palmsten K, Huybrechts KF, Mogun H, et al. Harnessing the Medicaid Analytic eXtract (MAX) to evaluate medications in pregnancy: design considerations. *PLoS One*. 2013; 8(6):e67405.
17. Desai RJ, Huybrechts KF, Hernandez-Diaz S, et al. Exposure to prescription opioid analgesics in utero and risk of neonatal abstinence syndrome: population based cohort study. *BMJ*. 2015;350:h2102.
18. Huybrechts KF, Bateman BT, Desai RJ, et al. Risk of neonatal drug withdrawal after intrauterine co-exposure to opioids and psychotropic medications: cohort study. *BMJ*. 2017;358:j3326.
19. Brown JS, Petronis KR, Bate A, et al. Drug adverse event detection in health plan data using the gamma Poisson shrinker and comparison to the tree-based scan statistic. *Pharmaceutics*. 2013;5(1):179–200.
20. Yih WK, Maro JC, Dashevsky I, et al. *Sentinel CBER Surveillance Protocol: Evaluation of HPV9 (Gardasil 9) Vaccine Safety Surveillance Using the TreeScan Data Mining Method*. Silver Spring, MD: Food and Drug Administration, US Department of Health and Human Services; 2016. https://www.sentinelinitiative.org/sites/default/files/vaccines-blood-biologics/assessments/Sentinel_PRISM_HPV9-Vaccine-Safety-Surveillance-Using-TreeScan.pdf. Accessed March 4, 2020.
21. Weston J, Bromley R, Jackson CF, et al. Monotherapy treatment of epilepsy in pregnancy: congenital malformation outcomes in the child. *Cochrane Database Syst Rev*. 2016; 11(11):CD010224.
22. Hernandez-Diaz S, Huybrechts KF, Desai RJ, et al. Topiramate use early in pregnancy and the risk of oral clefts: a pregnancy cohort study. *Neurology*. 2018; 90(4):e342–e351.
23. Hernández-Díaz S, Mitchell AA, Kelley KE, et al. Medications as a potential source of exposure to phthalates in the U.S. population. *Environ Health Perspect*. 2009;117(2): 185–189.
24. Kelley KE, Hernández-Díaz S, Chaplin EL, et al. Identification of phthalates in medications and dietary supplement formulations in the United States and Canada. *Environ Health Perspect*. 2012;120(3):379–384.