Data-Driven Queries between Medications and Spontaneous Preterm Birth Among 2.5 Million Pregnancies

Ivana Mari [1], **Virginia D. Winn**[2], **Evgeniya Borisenko**[3], **Kari A. Weber**[1], **Ronald J. Wong**[1], **Natali Aziz**[2], **Yair J. Blumenfeld**[2], **Yasser Y. El-Sayed**[2], **David K. Stevenson**[1], **Gary M. Shaw**[1]

[1]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA

[2]Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA, USA

[3]School of Engineering, Stanford University, Stanford, CA, USA

Abstract

Background: Our goal was to develop an approach that can systematically identify potential associations between medication prescribed in pregnancy and spontaneous preterm birth (sPTB) by mining large administrative "claims" databases containing hundreds of medications. One such association that we illustrate emerged with antiviral medications used for herpes treatment.

Methods: IBM MarketScan® databases (2007–2016) were used. A pregnancy cohort was established using ICD-9/10 codes. Multiple hypothesis testing and the Benjamini-Hochberg procedure that limited false discovery rate at 5% revealed, among 863 medications, 5 that showed odds ratios (ORs) of less than 1. The statistically strongest was an association between antivirals and sPTB that we illustrate as a real example of our approach, specifically for treatment of genital herpes (GH). Three groups of women were identified based on diagnosis of GH and treatment during the first 36 weeks of pregnancy: 1) GH without treatment; 2) GH treated with antivirals; 3) no GH or treatment.

Results: We identified 2,538,255 deliveries. 0.98% women had a diagnosis of GH. Among them, 60.0% received antiviral treatment. Women with treated GH had OR<1, (OR [95% CI] = 0.91 [0.85, 0.98]). In contrast, women with untreated GH had a small increased risk of sPTB (OR [95% CI] =1.22 [1.14, 1.32]).

Conclusions: Data-driven approaches can effectively generate new hypotheses on associations between medications and sPTB. This analysis led us to examine the association with GH treatment. While unknown confounders may impact these findings, our results indicate that women with untreated GH have a modest increased risk of sPTB.

CORRESPONDING AUTHOR: Ivana Mari , PhD**,** Research Scientist, Division of Neonatal and Developmental Medicine, Department of Pediatrics, Stanford University School of Medicine, 1265 Welch Rd, X1C55, Stanford, CA 94305, USA, ivanam@stanford.edu.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

For each medication, odds ratios (ORs) for risk of sPTB were computed by logistic regression. Multiple hypothesis testing was performed under the null hypothesis that a medication is not associated with sPTB. The Benjamini-Hochberg procedure was applied to control false discovery rate at 5% and to determine significant p-values.

INTRODUCTION

Use of medications in pregnancy has increased over the last few decades.(Ayad and Costantine, 2015; Mitchell and others, 2011) By 2008, 70% of women reported using at least one prescription medication during pregnancy, an increase from 49.4% during 1997–2003.(Mitchell and others, 2011) It has been estimated that the average number of prescriptions per pregnancy is 1.8.(Mitchell and others, 2011) With 3.95 million births in the US in 2016,(Martin and others, 2018) this implies more than 7 million medications being prescribed to pregnant women per year. The widespread use of medications warrants further investigation into potential associations with sPTB. Large administrative "claims" databases containing records on millions of patients can provide large enough sample sizes for studying various medication exposures to generate hypotheses and discover new potential associations with sPTB.

Others have recently described the importance and use of claims data to investigate pregnancy exposures and outcomes. Ailes et. al.[26] noted some of the attendant issues of using such data in their recent paper in this journal.

Our goal was to develop a method that can systematically identify potential associations between reported medication claims and, specifically, the adverse pregnancy outcome of spontaneous preterm birth (sPTB) by mining the IBM MarketScan® Commercial Claims and Encounters database. This database contains over 2.9 billion drug prescription claims, including time when prescriptions were processed and drug dosages. We employed computational, data-driven techniques that offer opportunities to perform simultaneous inferences about the use of multiple medications. Such methods have only recently been considered for discovery of unknown drug effects and interactions.(Boland and others, 2017; Chiang and others, 2018; Tatonetti and others, 2011; Tatonetti and others, 2012) Tatonetti et. al. successfully used data mining of adverse drug events (ADEs) to discover several previously unidentified drug-to-drug interactions.(Chiang and others, 2018; Tatonetti and others, 2012)

Our approach required the development of an algorithm to establish a pregnancy cohort with estimated gestational ages for this large database. The algorithm extends on previous related works by using ICD-10 codes and by addressing specific issues related to IBM MarketScan® databases. Here we describe the approach and highlight one of the potential associations that

emerged from multiple hypothesis testing, an inverse association between sPTB and antiviral drugs that are most commonly used for treatment of genital herpes.

METHODS

Determining Pregnancy Cohort

IBM MarketScan® Commercial Claims and Encounters databases that include health insurance claims data from across the US from 2007 to 2016 were used. A cohort of women aged 12–55 years with singleton pregnancies was identified from the IBM MarketScan® Inpatient Admissions database by using *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9) diagnosis codes (640–650, 652–679, V27.0, V27.1, V27.9; only codes ending in ***.*1 and ***.*2 were used), procedure codes (72.**, 73.22, 73.59, 73.6, 74.0–74.2, 74.4, 74.99), ICD-10 codes (O80-O83, O60.1*X0, O60.1*X1, O42.0 except O42.011, O42.1 except O42.111, O42.9 except O42.911) and diagnosis-group related (DRG) codes (774–768). To ensure that only singleton pregnancies were included, deliveries with codes for multi-fetal pregnancies (ICD-9 codes 651, V27.2-V27.7, V-91; ICD-10 codes O30) were excluded. To avoid dependencies for women with multiple pregnancies, only the first pregnancy recorded in the IBM MarketScan® databases was included. We excluded molar and ectopic pregnancies (ICD-9 codes 630 and 633; ICD-10 codes O00 and O01; DRG code 777), spontaneous and induced abortions (ICD-9 codes 632, 634–637; ICD-10 codes O03), complications following abortion, ectopic or molar pregnancies (ICD-9 code 639; ICD-10 codes O04 and O08), and abortions with/without D&C (DRG codes 770 and 779).

Continuous Enrollment Requirement

Only women who were continuously enrolled for at least nine months prior to admission for delivery were included.

Algorithm for Estimating Gestational Age (GA) at Delivery

IBM MarketScan® databases do not contain records for GA at delivery. To estimate GA at the delivery, we developed an algorithm based on codes for the number of completed weeks of gestation (ICD-9 codes 765.2, 766.21, 766.22; ICD-10 codes Z3A; Supplemental Table 1A) that we refer to as *GA codes,* and codes for preterm labor with preterm delivery (ICD-9 644.21; ICD-10 codes O60.1*X0) that we refer to as *sPTB codes.*

We identified the following groups of women:

1. Women with a GA code in their records. GA was assigned from the GA code.

2. Women without a GA code in their records, but with a GA code amongst their infant's' records. GA was assigned from their infant's' GA codes.

3. Women without a GA code. Among them, women with a sPTB code in their records, were assigned a GA of 34 weeks (which was the mean GA among women with sPTB in groups 1 and 2); otherwise, a GA of 39 weeks was assigned.

To obtain GAs from infant records (group 2), single liveborn infants were identified (ICD-9 codes V30 and V39; ICD-10 codes Z380-Z382). Mothers were then linked to their infant by using a unique family identifier that was introduced in the IBM MarketScan® databases in 2011. Infants of age>0 were excluded. To distinguish multiple infants born to the same mother, in addition to linking by the family identifier, we also linked a mother's record with her infant's record only if the admission date of the infant was between the mother's delivery admission and discharge dates.

For women in groups 1 and 2, we next checked whether they had a sPTB code. Women who had both a code for sPTB and a GA code for GA greater than 37 weeks were excluded (n = 2,663). Women who did not have a sPTB code, but had a GA code showing less than 37 weeks of gestation, were assumed to have a missing sPTB code and were included in the analyses. In the identified pregnancy cohort, 11.9% women had GA codes, either from their own or infant's record. Among women with sPTB, the frequency of GA codes was 28.7%, and the mean GA was 34.3 weeks. The algorithm is summarized in Supplementary Table 1B, which also shows the number of women at each stage of the algorithm.

Identifying sPTB

For women without a GA code in their records, the distinction of sPTB was made by codes for preterm labor with preterm delivery (ICD-9 code 644.21; ICD-10 codes O60.1*X0, O42.012).

For women with a GA code, sPTB was identified by a GA code <37 weeks accompanied by a code for the following diagnoses:

1. premature rupture of membranes (ICD-9 codes 658.1; ICD-10 codes O42.012, O42.013, O42.019, O42.112, O42.113, O42.119, O42.9 except O42.911);

2. early or threatened preterm labor (ICD-9 codes 644.*; ICD-10 codes O60.1*X0, O60.0).

Determining Delivery Dates

For women linked to their infants, their delivery dates were chosen to be equal to the admission date of their infants. For those not linked, the delivery date was chosen to equal a woman's delivery admission date.

Data-Driven Analysis of Medications in Pregnancy Cohort

We used the IBM MarketScan® Outpatient Pharmaceutical Claims database to link each woman in the established pregnancy cohort to all medications she received during pregnancy, starting at the estimated conception date. Since we were interested in a medication effect on preterm birth, we consider exposure during first 36 weeks of pregnancy, or until the delivery date, whichever occurred first. Women were divided into two groups based on whether they had sPTB or not. For each medication that appeared in the pregnancy cohort, unadjusted odds ratios (ORs) for the risk of sPTB were computed by univariate logistic regression. Specifically, for N medications that we identified were used in the pregnancy cohort, we first obtained N ORs for the risk of sPTB, denoted$(or_1, or_2, ..., or_N)$. For

each medication j we tested a null-hypothesis H_j that medication j was not related to sPTB, that is

$$H_j: or_j = 1 \quad j = 1, ..., N$$

(1)

resulting in N p-values, one for each hypothesis. Benjamini-Hochberg procedure(Benjamini and Hochberg, 1995) was then applied to control the false discovery rate (FDR) at 5% and to determine p-values that contradict the hypothesis. For comparison, we also performed the Bonferroni correction to control the family-wise error rate at 0.05. We observed that ORs > 1 were often a result of the indication for the medication and not necessarily the medication itself (*indication bias*). For example, we expect insulin to have high ORs as diabetes is a risk for sPTB. An approach to minimize such effects, would be to use propensity score matching[27] prior to generating hypotheses. Here, we initially focused on medications that appeared to have a "lower risk" for sPTB, that is have ORs < 1, with the reasoning that lowered risks identified through such hypothesis-free approach might not require propensity score matching. Out of 863 medications used during pregnancy, the Benjamini-Hochberg procedure yielded 5 that were associated with reduced risk of sPTB. For comparison, the Bonferroni correction, which is typically more conservative than Benjamini-Hochberg, was applied instead of Benjamini-Hochberg procedure, yielding one association with OR<1. Among the five medications chosen by Benjamini-Hochberg procedure, the medication with the smallest p-value, and the one chosen by Bonferroni correction was valacyclovir (OR=0.88), which is an antiviral medication commonly used for the treatment of genital herpes. We followed with a more detailed analysis into this association.

Identifying Patients with Genital Herpes

Our analysis, instead of focusing on valacyclovir only, was conducted more generally to include any of the three main antiviral medications commonly used to treat genital or oral herpes (valacyclovir, acyclovir or famciclovir) during first 36 weeks of pregnancy.

Within the established pregnancy cohort, three groups of women were identified based on their having a diagnosis of genital herpes (any, acute recurrent, etc.) and filling a prescription for antiviral medications (treatment) in first 36 weeks of pregnancy for term birth, or during the pregnancy in the case of preterm birth: Group 1) herpes and did not receive an antiviral medication, i.e., valacyclovir, acyclovir or famciclovir during the first 36 weeks of pregnancy (or prior to preterm birth); Group 2) herpes and received one of the three medications during the first 36 weeks of pregnancy (or prior to delivery in the case of preterm birth); and Group 3) the referent group of no herpes during pregnancy and did not receive an antiviral medication.

Practice guidelines for women with genital herpes in pregnancy recommend suppressive therapy for recurrent genital herpes (without symptoms) starting at 36 weeks of pregnancy. (2007) As a result, women with herpes who reach term are more likely to receive suppressive therapy, leading to a potential selection bias and an artificially lower risk of sPTB for women with genital herpes who received an antiviral treatment. Furthermore, any

treatment received after 36 weeks of pregnancy can have no impact to the event of sPTB, that would had occurred by then. For these reasons, women with herpes treated after 36 weeks of pregnancy (18.6% among women with herpes) were excluded from the treated group and were included in the group of women with untreated herpes in the first 36 weeks of pregnancy (Group 1). For the same reasons, we only considered women with their first listed diagnosis of genital herpes prior to 36 weeks (95.1% among women with herpes).

Women were identified as having a genital herpes diagnosis by diagnostic codes for genital herpes (ICD-9 codes 054.1; ICD-10 codes A60).

As we focused on genital herpes, we excluded women with diagnoses of any other herpes infection (ICD-9 codes 054.0, 054.2–054.9; ICD-10 codes B00), herpes zoster (ICD-9 codes 053; ICD-10 codes B02), or chicken pox (ICD-9 codes 052; ICD-10 codes B01) from the referent group.

RESULTS

Multiple Hypothesis Testing

The pregnancy cohort included 2,538,255 live, singleton births from 2007 to 2016. There were 149,551 (5.9%) women with sPTB. The prevalence of different medication classes in the established pregnancy cohort is shown in Figure 1. Among all women, 863 medications were prescribed and were considered in multiple hypothesis testing. By performing the Benjamini-Hochberg procedure to control the FDR at 5%, we identified 290 statistically significant p-values. Figure 2 shows p-values and their critical values for the Benjamini-Hochberg procedure. Among significant tests, there were only 5 with OR <1, including tests for valacyclovir. For comparison, the Bonferroni correction to control the family-wise error rate at 0.05, resulted in 165 significant p-values, out of which only valacyclovir had OR <1. To obtain the empirical distribution under the null hypothesis and confirm that z-values for p-values identified by the Benjamini-Hochberg procedure as significant were in the tail of the distribution,(Efron, 2004) the histogram of z-values for all 863 medications is shown in Figure 3. The z-values for significant p-values are shown in orange and are indeed in the tail of the distribution.

Case Example: Genital Herpes

Among the five medications identified to have an associated OR <1, valacyclovir is used to treat oral and genital herpes and herpes zoster or chicken pox. We explored this association further, by focusing on the patients with genital herpes. Genital herpes during pregnancy was diagnosed in 26,152 (1.0%) women and in 24,858 (0.98%) women in the first 36 weeks. Among the latter, 14,906 (60.0%) were treated with antiviral medications at any time in pregnancy. 13,542 (90.8% of all treated) women received the treatment in the first 36 weeks. The remaining 1,364 (9.2%) women treated after 36 weeks and women with untreated herpes (9,952 (40%)) constituted a group of women with untreated herpes before 37 weeks.

Table 1 shows the prevalence of sPTB for both women with herpes diagnoses and the referent group (no herpes, no treatment), along with the resulting ORs and confidence intervals (CIs). To examine the association between antiviral medication use and risk of

sPTB, we compared women with genital herpes – both untreated and treated women in the first 36 weeks of pregnancy – against the referent group. Women with untreated herpes in the first 36 weeks had a modestly higher risk of sPTB compared to the referent group (OR [95% CI] = 1.22 [1.14, 1.32]). In contrast, the treated herpes group did not, OR = 0.91 [0.85, 0.98].

We performed a sensitivity analysis in which we identified all women diagnosed with genital herpes at *any time* prior to pregnancy as well as during the first 36 weeks of pregnancy. By repeating the same analysis for this group, we obtained similar results as before (OR [95% CI] = 1.26 [1.20, 1.33] for untreated herpes and OR 95% CI] = 0.97 [0.91, 1.03] for treated herpes).

A second sensitivity analysis was performed where all women diagnosed with diabetes type I or II (ICD-9 codes 250, ICD-10 codes E10, E11), or gestational diabetes (ICD-9 codes 648.8, ICD-10 O24) were excluded from the cohort. The results were not affected (OR [95% CI] = 1.21 [1.11, 1.3] for untreated herpes and OR [95% CI] = 0.91 [0.85, 0.99] for treated herpes).

Finally, we performed a sensitivity analysis to investigate the impact of estimated gestational age. Specifically, women who delivered at term had genital herpes during first 36 weeks of pregnancy and received a treatment close to the 36th week of GA could be misclassified (treated vs. untreated) if the ICD9/10 code for GA is missing and the estimate of their GA is not precise. As explained in the Methods, women at term with unknown GA were assigned GA of 39 weeks (which equals the mean GA calculated from term women with known GA in the cohort). If this value would be increased to more than 39 weeks, more women with herpes would be classified as untreated during first 36 weeks therefore weakening the obtained results. And vice versa, it the assigned unknown GA would be decreased to less than 38 weeks, the effect observed in results would get stronger. We performed two additional analysis where the unknown assigned GA for term women was extended to: 1) 39 weeks and 3 days, resulting in OR [95% CI] = 1.1 [1.03, 1.19] for untreated herpes and OR [95% CI] = 0.94 [0.88, 1.02] for treated herpes 2) 40 weeks, resulting in OR [95% CI] = 1.04 [0.97, 1.12] for untreated herpes and OR [95% CI] = 1.01 [0.94, 1.09] for treated herpes.

DISCUSSION

We performed a discovery-based study to query a large number of medications available in IBM MarketScan® databases for their potential associations with sPTB. The discovery objective of this study was strengthened by its large sample size, providing enough power to perform hypothesis-free analysis for associations of various medications with sPTB.

Such a hypothesis-generating study was contingent upon establishing a pregnancy cohort. We presented a detailed algorithm that identified deliveries and estimated GA based on ICD9/10 codes for the number of completed weeks of gestation and codes for PTB. To circumvent the lack of GA data, one previous method assumed a fixed duration of pregnancy,(Johnsen and others, 2008; Raebel and others, 2005) an approach not applicable

when investigating PTB. ICD-9 codes specifying the number of completed weeks of gestation have been used previously for estimating GA.(Ailes and others, 2016; Li and others, 2013; Mines and others, 2014),[26]

While our algorithm uses the same ICD-9 codes employed in previous studies,(Ailes and others, 2016; Li and others, 2013; Mines and others, 2014),[26] it differs from those studies in several ways. First, inclusion of years 2015–2016 in the IBM MarketScan® databases provided us with the opportunity to utilize ICD-10 codes (implemented in 2015) that contain more precision in their definition of gestational weeks (there is a different code for each week of gestation). ICD-9 codes contain only two-week precision, except for 24 weeks of gestation (Supplemental Table 1A). Second, while in the Li et al. study,(Li and others, 2013) it was assumed that all term deliveries have a GA of 270 days, we determined GA for term deliveries more precisely, again using ICD-9/10 codes. While Ailes et al established the pregnancy cohort using IBM MarketScan databases (using ICD-9 codes), in their approach they did not link mothers to infants to obtain GA codes. However, these codes are more often contained in infant records than in mothers[28] and allowed us for a better estimate of GA. Finally, we found that the IBM MarketScan® database has occasional errors where a mother or her baby is assigned both codes for term and preterm delivery. We addressed these errors as explained in the Methods. Algorithms that estimate GA based on the presence/ absence of the preterm birth code or based on claims for routing prenatal screening tests that are taken at a narrow window of GA, have also been proposed and compared.(Margulis and others, 2013) Among the algorithms compared in that work, the best estimate was obtained by assuming a GA of 35 weeks for deliveries with PTB code, and a GA of 39 weeks for deliveries without the code. In our algorithm, to PTB deliveries for which exact GA was unknown we assigned GA of 34 weeks, as this was the mean value of GA for PTB deliveries for which GA was known in our cohort.

Our hypothesis-free analytical approach identified a common medication by both the Benjamini-Hochberg procedure and Bonferroni correction that has OR <1, i.e., valacyclovir, that is used for the treatment of oral and genital herpes, and herpes zoster or chicken pox. This hypothesis-generated result motivated us to explore in greater detail whether genital herpes treatment is associated with lower risk of sPTB.

Genital herpes is one of the most widespread sexually transmitted infections in the world. (Patel and others, 2017) In pregnancy, the estimated seroprevalence of genital herpes (HSV-2) is 22–25%,(Gardella and Brown, 2007; Xu and others, 2007) whereas primary infection during pregnancy is estimated to be 2–3%.(Li and others, 2014) Neonatal herpes caused by *in utero* infections from an exposed mother, while rare, can have devastating consequences for the infant.(2017) The impact that treatment of genital herpes can have on the risk of sPTB is not well established. A cohort study of 662,913 pregnant women with Kaiser Permanente insurance, demonstrated that untreated genital herpes in the first two trimesters may be associated with increased risk of PTB (OR [95% CI] = 2.23 [1.80, 2.76]). (Li and others, 2014) This study showed that treatment with an antiviral drug acyclovir almost eliminates the risk (OR [95% CI] = 1.11 [0.89, 1.38]). A double-blind randomized trial among 200 women with genital herpes at Mulago Hospital, Uganda showed a benefit of acyclovir treatment during 28–36 weeks of pregnancy, in reducing the incidence of

preterm birth.(Nakubulwa and others, 2017) Women were randomly assigned to one of the two equally-sized groups, one treated with acyclovir and the other given a placebo. Treated women had a reduced risk of preterm birth (RR [95% CI] = 0.41 [0.20, 0.85]). Our observations in this large claims database between treatment of genital herpes and the risk of spontaneous preterm birth are reasonably consistent with these published studies.(Li and others, 2014; Nakubulwa and others, 2017) Because the risk of ascending *in utero* infections from an exposed mother is low, the effect of treatment on spontaneous preterm birth is more likely to be through its effect on treating the maternal infection and the suppression of inflammatory signaling that might contribute to the initiation of early labor and delivery – and not the result of preventing infection of the fetus. Additionally, another hypothesis may question in antivirals used to treat genital herpes are actually treating another herpesviridae or other virus which has not yet been considered as a sPTB risk factor.

Use of claims data has several attendant limitations. One limitation in this analysis comes from lack of information in IBM MarketScan® databases about maternal characteristics including race/ethnicity, height, weight and smoking, each a possible confounder for the observed association. While it is well known that African-American women have a higher risk of preterm birth,(Goldenberg and others, 2008) we were not able to address this in the analysis due to lack of racial/ethnic information. Another limitation of the clinical data is that medication "exposure" designation is assumed on the basis of prescription fillings. We cannot know whether women actually took the medications prescribed. Finally, as IBM MarketScan® claims contain private health insurance claims, they do not typically represent the entire population. This also implies that among the women with untreated herpes, the reason for no treatment was not a lack of available coverage.

The observed prevalence of genital herpes in our cohort of 1.0% is significantly lower than what is observed in the general population (seroprevalence of 22%). Interestingly, the same was the case in a study (2017) that used a different database. Possible reasons could be missing diagnoses for patients who are asymptomatic and/or under-reporting. Additionally, this observed low prevalence may represent more primary genital herpes cases, which are generally easier to diagnose than recurrent cases. Hence, perhaps antiviral therapy has greater impact due to treatment of more severe primary infection.

In this data-driven screening for potential associations we screened for unadjusted ORs. Such an approach may be challenged by confounding influences due to underlying indication or to biases from prescriptions of concomitant medications – each of these potential biases may significantly alter unadjusted ORs. More generally, while screening methods to generate new hypotheses via multiple hypothesis testing or machine learning techniques may be effective at identifying potentially interesting associations, each generated hypothesis needs to be followed up by a close examination of indications, timing of treatment, and related diagnoses and medications, to minimize potential biases. Here, we illustrated some of these issues and how we addressed them in the genital herpes study.

Data mining as proposed here is a discovery-based approach. Its value lies in its ability to generate new hypotheses and identify potentially interesting association. A challenge with data-mining is that, the multiple hypothesis testing approach is typically expected to yield,

among hypothesis it finds to be significant, a subset of false discoveries (Tatonetti, 2019). For hypotheses generated by our presented data-mining approach, this may be subsequently overcome by performing follow-up studies to validate the findings using different datasets (e.g., Electronic Health Records, Optum database). Experimental studies, if possible to conduct for specific hypothesis, would provide further evidence.

Finally, our sensitivity analysis demonstrated the importance of knowing the GA to avoid misclassification in the case study of genital herpes. We observed that the association of either treated or untreated herpes with sPTB disappeared as the unknown GA for term deliveries was set to 40 weeks. We expect that GA of 40 weeks is an overestimate for three reasons: 1) it has been shown that the best estimate for unknown GA is 39 weeks[19]; 2) the mean GA for term deliveries with known GA was 39 in our cohort; 3) assuming 40 weeks for the GA estimate in the data-driven analysis identified medications used to assist conception, implying that the GA was too long. However, due to unknown exact GA for women with genital herpes who delivered at term, we cannot rule out a possibility that that the obtained results on genital herpes are due to misclassification. In fact, this case study, points out at the limitations of large administrative data bases. Further analysis in a data set with known GA is necessary.

In this large study, we performed simultaneous inferences on potential associations between medication use and risk of sPTB, for all medications that we identified during pregnancy. This led to a classic statistics problem of multiple hypothesis testing and identified 5 medications inversely associated with sPTB, including valacyclovir. The established pregnancy cohort and the proposed data-mining method can be used to analyze other adverse outcomes of pregnancy including birth defects. Given the sparsity of results related to the risk of sPTB and treatment of genital herpes (the main indication for these medications), we further analyzed this observed association. We observed that patients with untreated genital herpes have an increased risk of sPTB. While these results may encourage more careful diagnosis and treatment of genital herpes in pregnancy, we emphasize that, since the observed risk of sPTB with untreated herpes was modest, it is plausible that this risk is due to unaccounted confounding or misclassification that cannot be teased out using the IBM MarketScan® databases. More research needs to be conducted to further investigate potential protective effect of antivirals on the risk of spontaneous preterm birth.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

Data Availability Statement

IBM MarketScan Research Databases are available to purchase by Federal, non-profit, academic, pharmaceutical, and other researchers. Use of the data is contingent on completing a data use agreement and purchasing the data needed to support the study. More information about licensing the IBM MarketScan Research Databases is available at: https://www.ibm.com/us-en/marketplace/marketscan-research-databases.

This project did not require review by Stanford Institutional Review Board for Human Subjects.

REFERENCES

2007. ACOG Practice Bulletin. Clinical management guidelines for obstetrician-gynecologists. No. 82 June 2007. Management of herpes in pregnancy. Obstetrics and gynecology 109(6):1489–1498. [PubMed: 17569194]

2017. Fetal and Neonatal Brain Injury. 5 ed. Cambridge: Cambridge University Press.

Ailes EC, Simeone RM, Dawson AL, Petersen EE, Gilboa SM. 2016. Using insurance claims data to identify and estimate critical periods in pregnancy: An application to antidepressants. Birth defects research Part A, Clinical and molecular teratology 106(11):927–934. [PubMed: 27891779]

Ayad M, Costantine MM. 2015. Epidemiology of medications use in pregnancy. Seminars in perinatology 39(7):508–511. [PubMed: 26358804]

Benjamini Y, Hochberg Y. 1995. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. J Royal Statist Soc, Series B 57(1):289–300.

Boland MR, Polubriaginof F, Tatonetti NP. 2017. Development of A Machine Learning Algorithm to Classify Drugs Of Unknown Fetal Effect. Scientific Reports 7(1):12839. [PubMed: 28993650]

Chiang CW, Zhang P, Wang X, Wang L, Zhang S, Ning X, Shen L, Quinney SK, Li L. 2018. Translational High-Dimensional Drug Interaction Discovery and Validation Using Health Record Databases and Pharmacokinetics Models. Clinical pharmacology and therapeutics 103(2):287–295. [PubMed: 29052226]

Efron B 2004. Large-Scale Simultaneous Hypothesis Testing. Journal of the American Statistical Association 99(465):96–104.

Gardella C, Brown ZA. 2007. Managing genital herpes infections in pregnancy. Cleveland Clinic journal of medicine 74(3):217–224. [PubMed: 17375803]

Goldenberg RL, Culhane JF, Iams JD, Romero R. 2008. Epidemiology and causes of preterm birth. Lancet (London, England) 371(9606):75–84. [PubMed: 18177778]

Johnsen SL, Wilsgaard T, Rasmussen S, Hanson MA, Godfrey KM, Kiserud T. 2008. Fetal size in the second trimester is associated with the duration of pregnancy, small fetuses having longer pregnancies. BMC pregnancy and childbirth 8:25. [PubMed: 18627638]

Li DK, Raebel MA, Cheetham TC, Hansen C, Avalos L, Chen H, Davis R. 2014. Genital herpes and its treatment in relation to preterm delivery. American journal of epidemiology 180(11):1109–1117. [PubMed: 25392064]

Li Q, Andrade SE, Cooper WO, Davis RL, Dublin S, Hammad TA, Pawloski PA, Pinheiro SP, Raebel MA, Scott PE, Smith DH, Dashevsky I, Haffenreffer K, Johnson KE, Toh S. 2013. Validation of an algorithm to estimate gestational age in electronic health plan databases. Pharmacoepidemiology and drug safety 22(5):524–532. [PubMed: 23335117]

Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. 2013. Algorithms to estimate the beginning of pregnancy in administrative databases. Pharmacoepidemiology and drug safety 22(1):16–24. [PubMed: 22550030]

Martin JA, Hamilton BE, Osterman MJK, Driscoll AK, Drake P. 2018. Births: Final Data for 2016. National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System 67(1):1–55.

Mines D, Tennis P, Curkendall SM, Li DK, Peterson C, Andrews EB, Calingaert B, Chen H, Deshpande G, Esposito DB, Everage N, Holick CN, Meyer NM, Nkhoma ET, Quinn S, Rothman KJ, Chan KA. 2014. Topiramate use in pregnancy and the birth prevalence of oral clefts. Pharmacoepidemiology and drug safety 23(10):1017–1025. [PubMed: 24692316]

Mitchell AA, Gilboa SM, Werler MM, Kelley KE, Louik C, Hernandez-Diaz S. 2011. Medication use during pregnancy, with particular focus on prescription drugs: 1976–2008. American journal of obstetrics and gynecology 205(1):51.e51–58.

Nakubulwa S, Kaye DK, Bwanga F, Tumwesigye NM, Nakku-Joloba E, Mirembe F. 2017. Effect of suppressive acyclovir administered to HSV-2 positive mothers from week 28 to 36 weeks of pregnancy on adverse obstetric outcomes: a double-blind randomised placebo-controlled trial. Reproductive health 14(1):31. [PubMed: 28253893]

Patel R, Kennedy OJ, Clarke E, Geretti A, Nilsen A, Lautenschlager S, Green J, Donders G, van der Meijden W, Gomberg M, Moi H, Foley E. 2017. 2017 European guidelines for the management of genital herpes. International journal of STD & AIDS 28(14):1366–1379. [PubMed: 28836892]

Raebel MA, Ellis JL, Andrade SE. 2005. Evaluation of gestational age and admission date assumptions used to determine prenatal drug exposure from administrative data. Pharmacoepidemiology and drug safety 14(12):829–836. [PubMed: 15800957]

Tatonetti NP, Denny JC, Murphy SN, Fernald GH, Krishnan G, Castro V, Yue P, Tsao PS, Kohane I, Roden DM, Altman RB. 2011. Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. Clinical pharmacology and therapeutics 90(1):133–142. [PubMed: 21613990]

Tatonetti NP, Ye PP, Daneshjou R, Altman RB. 2012. Data-driven prediction of drug effects and interactions. Science translational medicine 4(125):125ra131.

Tatonetti NP. 2019. The next generation of drug safety science: coupling detection, corroboration, and validation to discover novel drug effects and drug-drug interactions. Clin Pharmacol Ther. 2018 Feb; 103(2): 177–179.

Xu F, Markowitz LE, Gottlieb SL, Berman SM. 2007. Seroprevalence of herpes simplex virus types 1 and 2 in pregnant women in the United States. American journal of obstetrics and gynecology 196(1):43.e41–46.

Prevalence of Prescriptions in Pregnancy
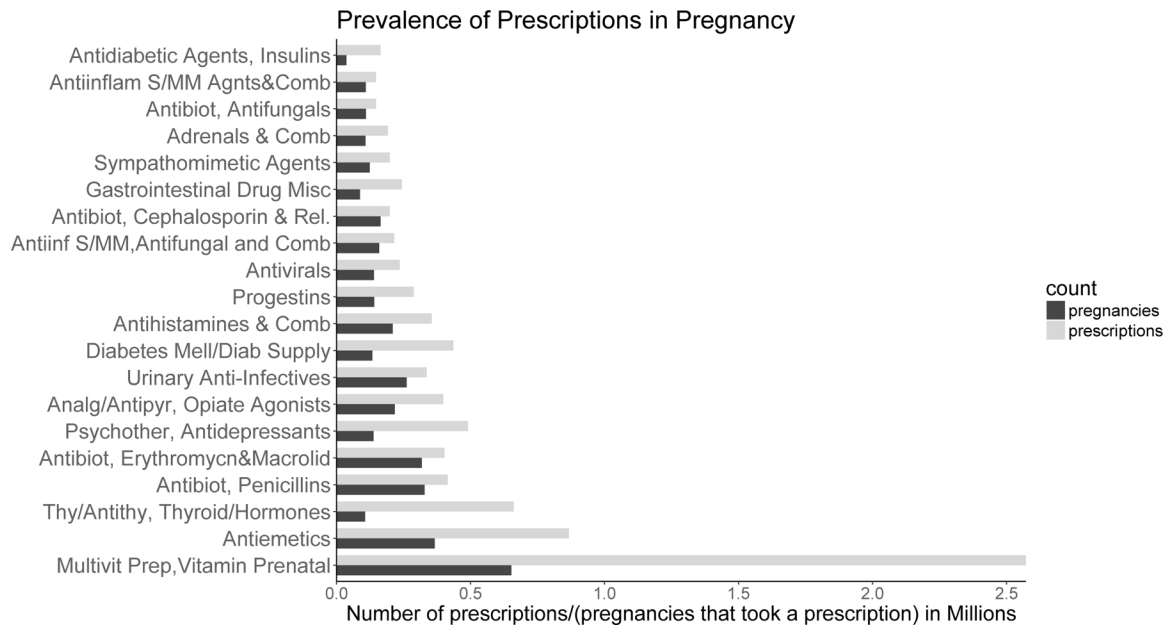
Figure 1.

Number of prescriptions filled in pregnancy (gray bars) and number of pregnancies that filled a precription (black bars) for different medication classes. 20 most frequent classes are shown.
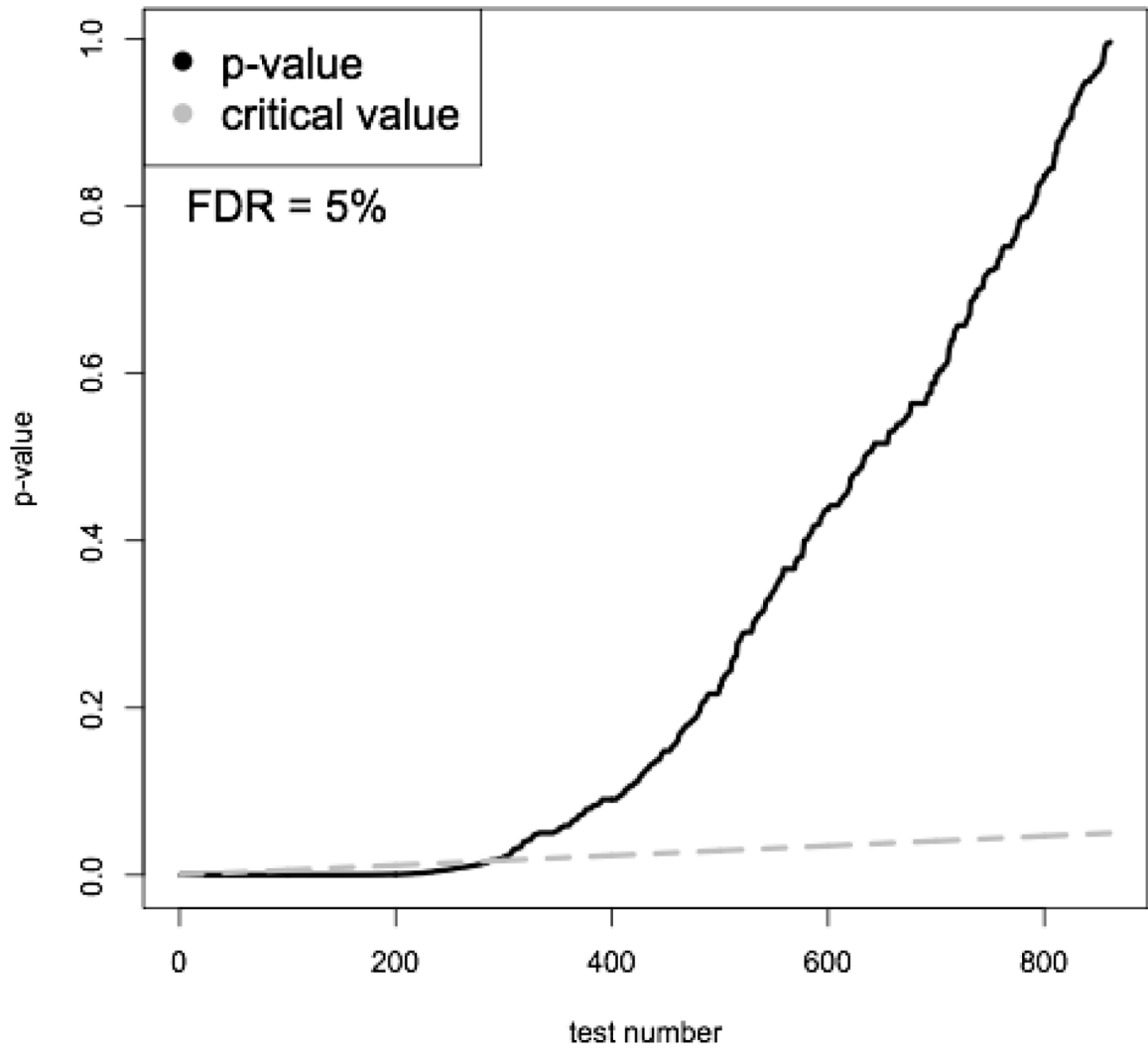
Figure 2.
p-values and critical values obtained using the Benjamini-Hochberg procedure.
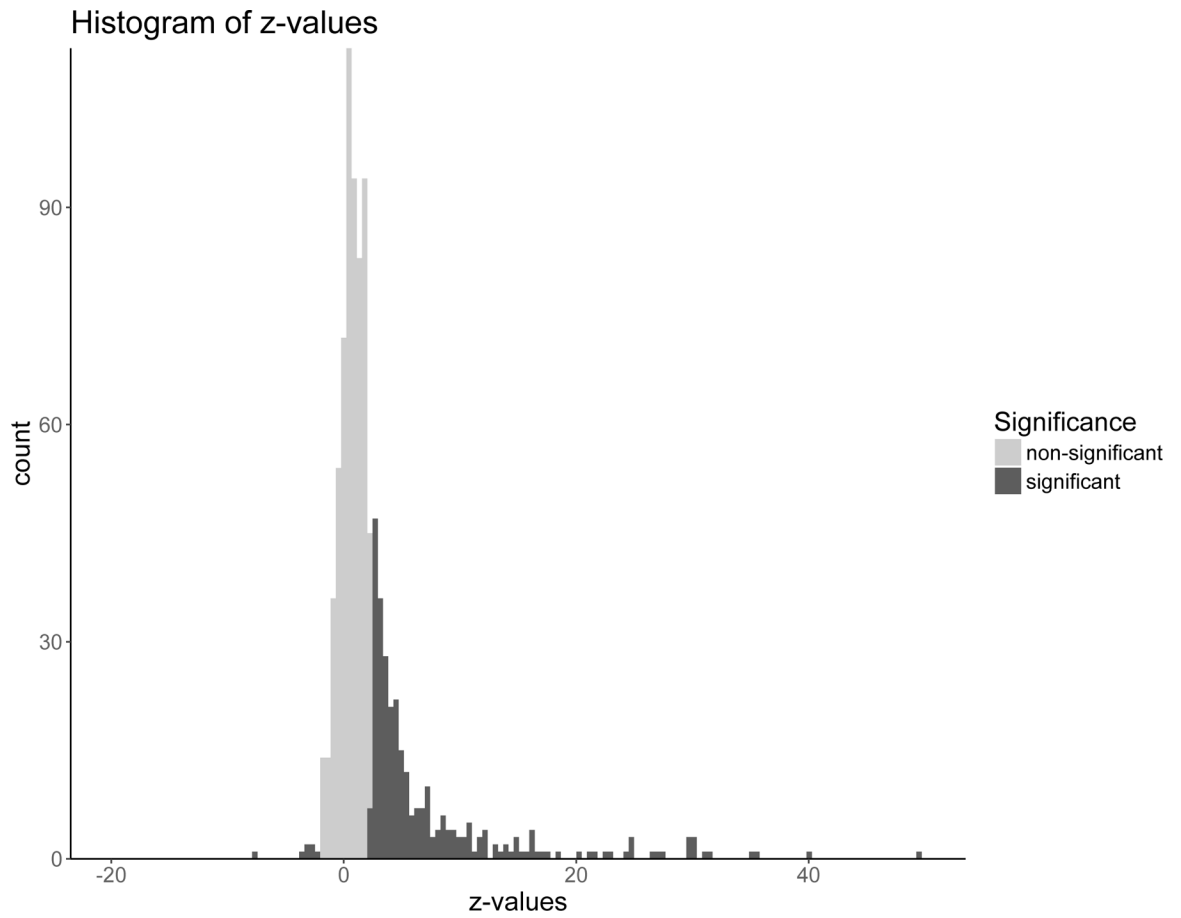
Figure 3.
Histogram of 863 z-values, one for each medication, obtained by testing multiple hypotheses.

Table 1.

Odds ratios for an association between treated and untreated herpes infection with spontaneous preterm birth.

		sPTB	%	Total	Compared with Referent group		Compared with Untreated Herpes	
					OR		OR	95% CI
Referent group		143,099	5.9	2,404,648			Referent	
Herpes	Treated	741	5.5	13,542	0.91	[0.85, 0.98]	0.74	[0.67, 0.82]
	Untreated	816	7.2	11,316	1.22	[1.14, 1.32]		