

Medication history-wide association studies for pharmacovigilance of pregnant patients

Anup P. Challa^{1,2,3}, Xinnan Niu⁴, Etoi A. Garrison⁵, Sara L. Van Driest^{6,7}, Lisa M. Bastarache⁴, Ethan S. Lippmann², Robert R. Lavieri¹, Jeffery A. Goldstein⁸ & David M. Aronoff^{5,7,9,10}

Abstract

Background Systematic exclusion of pregnant people from interventional clinical trials has created a public health emergency for millions of patients through a dearth of robust safety data for common drugs.

Methods We harnessed an enterprise collection of 2.8 M electronic health records (EHRs) from routine care, leveraging data linkages between mothers and their babies to detect drug safety signals in this population at full scale. Our mixed-methods signal detection approach stimulates new hypotheses for post-marketing surveillance agnostically of both drugs and diseases—by identifying 1,054 drugs historically prescribed to pregnant patients; developing a quantitative, medication history-wide association study; and integrating a qualitative evidence synthesis platform using expert clinician review for integration of biomedical specificity—to test the effects of maternal exposure to diverse drugs on the incidence of neurodevelopmental defects in their children.

Results We replicated known teratogenic risks and existing knowledge on drug structure-related teratogenicity; we also highlight 5 common drug classes for which we believe this work warrants updated assessment of their safety.

Conclusion Here, we present roots of an agile framework to guide enhanced medication regulations, as well as the ontological and analytical limitations that currently restrict the integration of real-world data into drug safety management during pregnancy. This research is not a replacement for inclusion of pregnant people in prospective clinical studies, but it presents a tractable team science approach to evaluating the utility of EHRs for new regulatory review programs—towards improving the delicate equipoise of accuracy and ethics in assessing drug safety in pregnancy.

Plain language summary

The exclusion of pregnant people during clinical drug development limits our understanding of drug safety in pregnancy. However, given that many approved medications are prescribed during pregnancy, we studied large sets of data from patient electronic medical records, including mother-baby pairs, to develop these drug safety insights. From 2.8 M medical records, we identified 1,054 drugs prescribed to pregnant patients. Combining computerized analysis with expert clinical review, we confirmed signals previously associated with neurodevelopmental defects in children born to drug-exposed mothers. We subsequently identified 5 commonly prescribed drug classes for which we believe this work warrants updated assessment of their safety. Our approach is not a replacement for inclusion of pregnant people in clinical trials but presents a new application of existing medical record data to improve the assessment of drug safety in pregnancy.

¹Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN 37203, USA. ²Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN 37212, USA. ³Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA. ⁴Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37203, USA. ⁵Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN 37203, USA. ⁶Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN 37232, USA. ⁷Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37203, USA. ⁸Department of Pathology, Northwestern University, Chicago, IL 60611, USA. ⁹Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37203, USA. ¹⁰Present address: Department of Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA. ✉email: sb108vg@gmail.com; aronoff@iu.edu

At the point of care, pregnant patients are a complex population: physicians must exercise caution in prescribing many common drugs to these patients, given the risks of toxicity for their developing fetuses¹. However, consideration of fetal toxicity in drug development is largely irregular. While teratogenicity scores established by regulatory agencies like United States Food and Drug Administration (FDA) are discrete, these criteria provide little concrete distinction among score classes, making it difficult for drug developers to accurately gauge the fetal toxicity risks of a molecule². FDA's updated teratology assessment guidelines in the 2014 Pregnancy and Lactation Labeling Rule aimed to increase the contextual relevance of developmental toxicity evaluation, but this guidance has been slow to translate to evaluative change at the point of care, which remains largely aligned with the previous five-pronged letter scale^{3,4}. The result is a vicious cycle that promotes the approval of drugs without adequate data on their safety and efficacy in pregnant populations, as expectant patients are routinely excluded from clinical trials, out of concern for fetal harm upon exposure to drugs with uncertain, pre-clinical teratogenicity data. In fact, of 213 new drugs approved by FDA between 2003 and 2012, only 5% contained human data in the pregnancy section of their labels⁵. These factors have created a substantial gap in knowledge on pharmacotherapy for diseases during pregnancy, restricting the number of treatments available to this population through insufficient data on the pharmacodynamics and pharmacokinetics (PK) of many maternal medication exposures. At the bedside, the result is undertreatment of chronic and acute illnesses in pregnant people from obstetricians' cautious fears of causing harm to their patients, alongside the increased risk of harm to fetuses from necessary prescriptions⁶.

While the ethics of excluding pregnant people from randomized, controlled drug trials (RCTs) remain in debate⁷⁻⁹, the ongoing unavailability of relevant drug safety and efficacy information underscores an urgent need for new methods to rapidly assess this information, to improve the quality of care for these underserved patients, and to ensure health equity for this complex population through contemporaneous drug labeling and marketing efforts. Such an opportunity for the discovery of drug safety insights for pregnant patients may be available through strategic analysis of large numbers of existing healthcare documents like electronic health records (EHRs) that were collected during routine patient care. Collectively, EHRs can uniquely replicate the natural history of pregnancy by linking medical information of pregnant patients and their neonates, such as mothers' prescriptions (while expectant) and the perinatal diseases of their children¹⁰⁻¹². This information allows for the creation of a unique framework of relational knowledge generation. Namely, EHR data may be stratified into distinct cohorts by patients' documented exposure—or lack thereof—to a drug of interest, facilitating the development of an inferential model to relate incidences of maternal drug exposure and neonatal disease¹¹. While these experiments are not a replacement for prospective safety data generation through the inclusion of pregnant people in clinical trials, the above platform of safety signal detection presents an ethical way of studying the effects of drug exposure in pregnant people with human data, on a significant scale and across all drug classes.

Existing literature that describes the safety of most drugs potentially prescribable in pregnancy remains overwhelmed by conflicting studies—the majority of which only present results from pre-clinical animal models of drug testing and the minority of which are empirical case reports or case series among relatively few patients¹³. Deciding to prescribe a drug to a pregnant patient involves balanced evaluation of the patient's need for treatment (drug efficacy) and the risk of injury to the patient's fetus (drug

safety). However, providers cannot make these informed decisions without robust and definitive safety data.

Previous work that has attempted to clarify knowledge on drug safety in pregnant patients has relied on observational and retrospective analyses of databases like public insurance claims, measuring the significance in the coincidence of a neonatal disease of interest and prescription of a drug of interest to the neonates' mothers^{10,14}. While these studies have added new—and often valuable—narratives of drug safety to the literature, our research is innovative because it uses EHR data, attempts relational inference, and probes such drug-disease relationships at scale. Collectively, these factors allow us to advance the ontological reliability and epistemological robustness of data-driven studies of adverse pregnancy outcomes¹¹.

Our research makes use of a database of 2.8 M EHRs at Vanderbilt University Medical Center (VUMC) to curate our experimental cohorts. The data innovation in studying EHR data over evaluating public insurance claims is that this choice mitigates significant demographic biases (e.g., poverty) that are present within public payor records. Overcoming the effects of such potentially confounding variables requires the integration of advanced methods of propensity scoring (PS) to properly evaluate the coincidence of maternal drug exposure and pediatric disease, which defines the key algorithmic design principle of parsimony and results in poor model performance¹⁵. In contrast, VUMC is an urban medical center that features a demographically diverse patient population, as previous studies using these EHR data affirm¹⁶. Indeed, self-reporting patient registries—another popular choice for observational data to study health outcomes in pregnancy—are also inherently limited in their integrity, as patients are often unreliable historians of their own care¹⁷. In contrast, our study promotes data integrity by studying provider-maintained healthcare information.

Technical innovation in this project also rests within the rigor of the analytical methods we employ¹¹. We apply a mode of systematic, relational inference to maternal drug exposure and perinatal disease that we believe is more directly and appropriately aligned with the etiology of drug-associated birth defects, compared to the highly coincidental frameworks that dominate the literature. We achieved inference suggestive of causality through harmonizing the validated phenome-wide association (PheWAS), which was originally developed at VUMC to discover genetic links to clinical phenotypes, with a rigorous, standardized consensus prioritization approach that considered clinical practice and RCT data to move from data-based associations towards etiology discovery¹⁸. By developing a medication history-wide association study (MedWAS) to suggest pharmacological determinants of neonatal diseases, we optimized on algorithms that underlie PheWAS to explore nascent patterns across the drug-disease hypotheses that our model revealed. In this way, we used MedWAS as a method of safety signal detection and management, approaching the design a target trial¹¹. Target trials are an epidemiological method of retrospective data analysis that make use of existing clinical information and high-powered statistical algorithms to create artificial subject profiles from all relevant and available patient data within a cohort of interest. This curation then allows for relational analysis of subjects' drug histories against a morbidity of interest, facilitating potential simulation of a clinical trial when prospective experiments are not feasible¹⁹⁻²¹. The approach in this manuscript alludes to a target trial by following similar approaches to data curation and stratification, statistical inference, and outcomes prioritization, though unlike the archetypal target trial developed by Hernán and Robbins for claims data and consortial data banks¹⁹, our distributed workflow relies on a single health system's mother-baby EHRs, meaning that some aspects of our procedure rely on manual evidence

synthesis, rather than harnessing end-to-end automation. Furthermore, our approach operates on known patterns of prescriptive behavior in pregnancy to determine treatment-exposed and non-exposed (i.e., “control”) cohorts in our data, providing a very limited basis to claim RCT-like randomization that is naturally resultant from a poor recapitulation of the many reasons why clinicians decide on specific treatments for their pregnant patients, within structured EHR data. Like the target trial, our research does not seek to replace the RCT. Nonetheless, to our knowledge, there have been very few (and relatively small) attempts at EHR-derived safety signal detection evaluating pregnant patients²², allowing us to innovate in exploring the power of this approach at scale^{11,23}.

Using MedWAS, we present systematic safety signal detection across all drugs prescribed to pregnant people and all diseases within neonatal EHRs at VUMC: herein lies the conceptual innovation of our approach. Historically, researchers studying the safety of pharmacotherapy in pregnancy with statistical methods have communicated through a “one drug—one disease—one publication” model. While this practice provides bandwidth for deep interrogation of a single drug-disease hypothesis, it further diversifies the pool of existing data that remains conflicting and inconsistent, since the methods in such papers can become overfitted for studying the safety of other drugs that are prescribable in pregnancy. In contrast, our approach is sufficiently reproducible to analyze maternal prescriptions and neonatal diseases across a large healthcare enterprise. We are unaware of such a drug-agnostic and phenotype-agnostic model in the available literature on drug safety in pregnancy.

We have a record of work in using statistical methods like PheWAS to generate strong hypotheses of efficacy for new drug development^{12,24,25}. Here, we apply that expertise to construct MedWAS as an innovatively scalable approach for the surveillance of drug safety in pregnancy. We also present potential avenues for complementarity between MedWAS and our previous attempts to develop a machine learning (ML) approach capable of identifying chemical structures that predispose drugs towards increased teratogenic risk when prescribed during pregnancy²⁶.

In this study, we identify 1,054 drugs historically prescribed to pregnant patients and develop a quantitative, medication history-wide association study. We integrate a qualitative evidence synthesis platform using expert clinician review for inclusion of biomedical specificity—to test the effects of maternal exposure to diverse drugs on the incidence of neurodevelopmental defects in their children. Not only do the results replicate known teratogenic risks and existing knowledge on drug structure-related teratogenicity; they also highlight 5 common drug classes for which we believe this work warrants updated assessment of their safety. This research is not a replacement for the inclusion of pregnant people in prospective clinical studies, but presents a tractable team science approach to evaluating the utility of EHRs for new regulatory review programs—towards improving the delicate equipoise of accuracy and ethics in assessing drug safety in pregnancy.

Methods

The approach that we describe below is an explanatory summary of the data preprocessing (for cohort selection) and informatics procedures (for drug-disease testing) that we provide in cookbook format in the “Supplementary Information” accompanying this manuscript, supporting Supplementary Tables 1–5 in the component “Supplementary Methods” section. A diversity and inclusion report for the maternal and neonatal EHRs we analyzed is also included as Supplementary Table 6 in the “Supplementary Discussion” section.

We tested the hypothesis that MedWAS can effectively establish relational inference between mothers’ exposures to drugs with uncertain safety and perinatal diseases in their neonates. In establishing the feasibility of our tool to accomplish post-market drug surveillance, we restricted ourselves to the analysis of only neurological morbidities as a base case, given that the ontologies that codify these diseases have strong bases of relational logic²⁷. We expect the general framework of the analytical and signal evaluation procedures we present here will be analogously applicable to the interrogation of neonatal diseases in other organ systems.

Ethical review. The Institutional Review Board (IRB) of Vanderbilt University approved the research and deemed that it was exempt from ethical approval and informed consent since it was not deemed to involve human subjects (IRB #191553), given its retrospective, observational nature and use of data collected during routine patient care.

Cohort selection. To mimic the enrollment of pregnant patients in a drug safety experiment, we used ML to curate and block appropriate treatment and control (drug-exposed vs. not drug-exposed) cohorts across all 1,054 agents that are documented as prescriptions to pregnant patients in eStar, VUMC’s EHR system. A listing of these agents is available as described in Supplementary Methods, supporting Supplementary Data 1. To select our cohorts, we probed VUMC’s Research Derivative (RD), a database of fully identified clinical and administrative information from 2.8 M patients that contains data like International Classification of Disease-9/10 (ICD-9/10) billing codes (which codify nearly all existing human morbidities), patient demographics, lab results, medications, and clinical narratives from five different relational health information systems that source directly from patient care^{28,29}. To effectively create experimental cohorts across the agents we probed from these data, we first established the following phenotyping rule as inclusion criteria for patient “enrollment” in treatment and control groups:

Population: RD; Include: Mom/baby link (1 or more), where specified medication (1 or more where date during mother EHR pregnancy=yes) and clinic note in baby EHR suggests record of care (1 or more postpartum).

Herein, our criteria for allocating pregnant patients to a drug treatment group required baseline, confirmed pregnancy among all candidate mothers, with a record of at least one prescription of the specified drug in the mother’s EHR during their entire gestational period and live-born delivery of a neonate who received their own EHR at VUMC (so their health outcomes were available for our analysis). Defining pregnancy and gestational period in a systematic way from the EHR remains a non-standardized analytical practice and therefore required us to develop an inferential approach reliant on a data dictionary of relevant ICD-10 codes for gestational period. For interested readers, we describe this approach in Supplementary Methods, across Supplementary Tables 1–5. We designed our inclusion criteria to maximize the data available to our model, so we could achieve the highest power for demonstrating preliminary proof of concept for our approach. Herein, we harnessed downstream evidence synthesis to vet our outcome associations, rather than establishing very tight inclusion (and exclusion) criteria a priori to mitigate confounders.

Data curation. Next, we leveraged a suite of natural language processing (NLP) tools to extract phenotypic attributes and maternal drug exposures from narrative EHR data among all patients within the 94,872 EHRs (48,434 mother-baby EHR pairs) who met our inclusion criteria for at least one study drug. These tools

included a general-purpose NLP tool (the 2015-indexed version of KnowledgeMap concept identifier (KMCI)^{30,31}, available through <https://www.vumc.org/cpm/cpm-blog/kmci-knowledgemap-concept-indexer>), ML-based clinical-note section tagger (the 2010-indexed version of SecTag^{32,33}, available to download at <https://www.vumc.org/cpm/cpm-blog/sectag-tagging-clinical-note-section-headers>), and version 1.3 of MedEx, an NLP algorithm for identifying medication exposures within free clinical text^{32,34} (available to download at <https://sbmi.uth.edu/ccb/resources/medex.htm>). KMCI identifies Unified Medical Language System concepts³⁵ using a shallow parser, word sense disambiguation, and semantic regularization, and includes a module to identify negation³⁰. MedEx uses context-free grammar and a rule-based approach to extract detailed medication information (including dose, frequency, and route) from free text. MedEx encodes an ingredient barcode for all drugs, such that drug mentions extracted from EHRs are continuously linked to existing drug ontologies from which additional pharmacological data may be mined (e.g., RxNorm concept unique identifier³⁶)^{32,34}. These standardized systems have been used to process more than 60 million documents at Vanderbilt and elsewhere. Here, we used them to capture all drug mentions and available ICD-9/10 codes and to facilitate requisite matching of free-text disease terms to concept unique identifiers for candidate mothers and their linked neonates, as well as to extract all available demographic information for “enrolled” mothers and babies. Enacted across all combinations of diseases and maternal drug histories in our population, our workflow enabled the curation and stratification of patient data to empower >1.7 M combinatorial drug-disease association experiments, as we describe below.

Implementation of MedWAS. PheWAS is a common, systematic ML approach to unearth associations between disease and genetic variants and to discover pleiotropy using EHR data linked to DNA. It is a method that scans phenomic data for genetic associations using Phecodes mapped to ICD-9/10 codes from the EHR. Multiple publications demonstrate that PheWAS is a feasible method to rapidly generate hypotheses on the underpinnings of disease^{18,37–40}. We repurposed the PheWAS framework to develop an innovative MedWAS, in identifying the extent to which the perinatal phenotypes in our cohorts are plausibly related to exposure to the drugs in each simulated safety experiment’s treatment group. Herein, our proof-of-concept MedWAS model took an input of babies’ neurological diseases from all mother-baby cohorts we constructed and outputted the maternal medication exposures putatively related to babies’ phenotypes. While it is easiest to envision our platform through the canonical stratification of mother-baby cohorts by maternal drug exposure, our adoption of neonatal disease-contingent inference across treatment-defined maternal cohorts allowed us to develop capacity for discovery of multiple drug exposures as etiologies for our phenotypes of interest.

MedWAS operated in direct analogy to PheWAS by using its component logistic classification methods (logit) to identify neonatal disease as a function of maternal exposure to a drug of interest and by reporting a *p*-value for each of these drug-disease tests that reflected the strength of logit alignment after correction for multiple testing of a drug across all neonatal diseases in our cohorts. In doing this across 1,054 native maternal drug exposures and the neurological subset of 1,678 EHR-embedded phenotypes—first, on a pilot-scale, with 5.7 K EHR pairs, and subsequently on our full data set of 49 K mother-baby dyads—each experiment was controlled by cases of neonatal disease linked to pregnant patients without a record of exposure to the test drug. Herein, we also computed an odds ratio (OR) as a proxy for the effect size of hypothetical drug-disease enrichment across each of our tested case and control populations. Because

there are known associations among the representations of input and output data and PheWAS model performance^{38–40}, we iteratively assessed MedWAS performance with several standard representations of the drug and disease data (i.e., different levels of Anatomical Therapeutic Chemical (ATC) codes for drug entities⁴¹ and Phecodes and ICD-9/10 codes for diseases⁴²) from our cohorts to prevent confounding of our results by data type. The list of 1,678 Phecodes we employed is publicly accessible through the open-source code for version 0.12.3 of the PheWAS package (see <https://github.com/PheWAS/PheWAS>).

Hypothesis Prioritization. While the explicit goal of our work was to establish a platform for generating hypotheses of drug safety that may be pursued in more targeted studies in the future, we affirm that a non-deterministic challenge in pursuing our experiments was accurate prioritization of MedWAS’s predicted drug-disease relationships by their clinical, biological, and statistical plausibility, given the number of association tests we executed rapidly within our analytical framework. We attempted to meet this challenge by ranking our results with the following heuristics: concordance with known fetal safety risks from published drug labels, a soft constraint of Bonferroni significance (with correction from baseline $p \leq 0.05$) and $OR > 1$, compelling clinical reviews from obstetrician and pediatrician consults on the plausibility of substantially implicated drug prescriptions and teratogenic outcomes, reproducibility between MedWAS outputs and the results from our previous work that identified drug structures linked to adverse birth outcomes²⁶, and evidence against “confounding by indication” from harmonizing systematic chart review of mothers’ baseline disease states with knowledge of known vertical disease transmission risks within our treatment cohorts. Our application of the *p*-value as a soft prioritization constraint that complemented systematic review from our clinical stakeholders aligns with guidance to this effect from American Statistical Association⁴³.

To parse MedWAS results we believed were not clinically plausible or were potentially confounded, we began by restricting all signals associated to nutraceutical products, as we recognized that patient history-informed capture of food and nutritional supplement use data in the EHR is highly unreliable. These agents are available over-the-counter (OTC) and often incompletely reported by patients, such that mention of the agent does not always imply true exposure during gestation⁴⁴.

Pediatrics evidence synthesis. Then, we consulted a pediatrician with expertise in clinical pharmacology on our study team to identify neurological Phecodes with unlikely manifestation in the perinatal period; these diseases were mainly neurocognitive (e.g., dyslexia) and therefore excluded from consideration as true model results. Our pediatrics consult further stratified higher-level versions of the phenotype embeddings in our model outcomes as incident in infants, toddlers, school-age children, or adolescents, based on disease pattern presentations from clinical practice. Consequently, we excluded all outcomes not plausibly detectable in infants.

Obstetrics evidence synthesis. Following our pediatrician’s review, we consulted a practicing obstetrician on our study team, who has training in clinical pharmacology and maternal-fetal medicine, to identify the plausibility of prescription of the drugs implicated in our model during pregnancy. In completing this review, our obstetrics consultant synthesized knowledge from her own prescriptive practice, prescriptive guidelines from American College of Obstetricians and Gynecologists, Society for Maternal-Fetal Medicine, departmental practice guidelines at Vanderbilt, and clinical decision software (CDS) like UpToDate⁴ and

Reprotox⁴⁵ to stratify our signals as “high-yield” and “low-yield” outcomes. We defined high-yield outcomes as those which demonstrated statistical significance, at least 1% coincidence rate between drug prescription and pediatric disease (such that, with our sample sizes of mothers prescribed each drug and neonates born with each disease, we prioritized only non-unary outcomes), and unclear prescriptive recommendations and/or practice guidelines for implicated drugs (e.g., FDA score C and conflicting case reports described in CDS). These drugs also had plausible prescription during the first trimester of pregnancy, when most neurological organ development occurs. Low-yield outcomes included signals rooted in drugs available OTC, such that EHR data on drug use were not reliable for our first-pass analysis, and signals with drugs sparsely prescribed to pregnant patients in the United States of America due to lack of regional drug supply and/or existing guidance against prescription of these drugs during pregnancy. Our consideration of the latter revealed to us that our low-yield signals may be artifactual noise from our inferential approach to defining gestational period, if these drugs appeared in pregnant patients’ EHRs before discontinuation, when providers first learned of their patients’ pregnancies.

Our designation of the yields of our signals was powered by a spreadsheet model we developed, which codified the considerations above by fields including the following: (1) “drug’s original indication” (to help identify potential cases of confounding by maternal morbidity—by which a neonate could inherit the mother’s disease or the drug’s associated adverse outcomes could be sequelae of pre-term birth precipitated by the disease for which the mother is treated); (2) “FDA drug class”; (3) “trimester of prescription”; (4) “intrapartum or immediate postpartum prescription?” (a response of “yes” to this question resulted in a signal’s relative de-prioritization, given our interest in antepartum exposures and the difficulty of perfectly ascertaining gestational period within the EHR); (5) “duration of prescription”.

In an *ad hoc* fashion, both consultants, as well as a pharmacologist, removed drugs from consideration which presented with implausible PK for their associated toxicities (e.g., non-systematic absorption).

Figure 1 provides a summary of our process for developing and vetting MedWAS data.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Results

We present MedWAS as a customizable process of generating hypotheses for post-market drug surveillance of drug safety in pregnancy, which takes strategic advantage of the milieu of primary medical care for pregnant patients and the data routinely generated through these encounters. We present key results from our platform below, along with a discussion of the advantages, several limitations, and positive reception of our attempt, which we believe collectively define opportunities for expansion of our approach as a systematic attempt at drug safety assurance that is powered by real-world evidence (RWE).

Proof of concept. *Prima facie*, we consider MedWAS successful for its robust capacity to accommodate the largescale testing that we envisioned: following our experimental design, pilot testing, and localized sensitivity analyses, we executed 1,770,290 drug-disease experiments using a high-performance cluster with 2,400 processor cores hosted by the Southern Crossroads server⁴⁶ for supercomputing.

As we describe in “Methods,” facing an abundance of generated data, we restricted analysis of the reliability of our results to a single physiological system, to allow for deep contextual analysis. Accordingly, we selected to analyze 1,414 neuroteratogenic signals meeting our aforementioned definition of statistical significance, given expertise in neuropathogenesis within our study team and the spatially and temporally focal nature of many neurodevelopmental anomalies to neurulation⁴⁷, which occurs in the first trimester of pregnancy⁴⁸. In analyzing this functional area, we assume that our insights are sufficiently generalizable to similar physiology in other organ systems, but we also acknowledge that signals among systemic developmental phenotypes may require more formal network analyses. In considering the validation procedures we describe in “Methods” and the evidence requirements we discuss below for signal confirmation, we found that MedWAS performed best on the

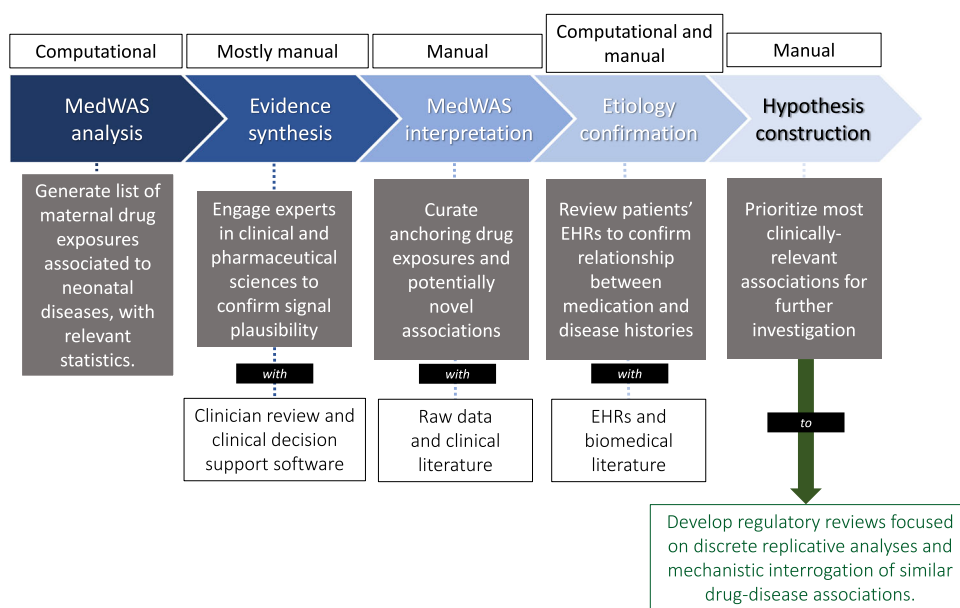


Fig. 1 Summary of the process to develop and assure the quality of MedWAS signals—the goal of our approach is to unearth and prioritize drug safety risks, towards generation of the highest quality hypotheses to inform new regulatory review programs. Engagement of obstetric, pharmacological, and regulatory stakeholders is inherent to this process.

Table 1 Example MedWAS Outcome^a: An example MedWAS outcome for a known teratogenic relationship (fetal phenytoin intoxication and chorea) shows statistical significance across several mother-baby pairs, as we expected.

| Drug | Disease | <i>p</i> | OR | # Disease + | % Disease + with Drug Exposure |
|-----------|--------------------------------|--------------------|------|-------------|--------------------------------|
| Phenytoin | Abnormal involuntary movements | 1×10^{-6} | 1.03 | 195 | 1 |

^aPer the “Data Availability Statement,” row-level positive control data may be available upon request, and the accompanying Supplementary Data 1 contains exposure counts for all drugs we studied. A collection of 21 similar results across 10 known teratogens supports our claim of proof of concept for our approach.

bases of ATC-4 and Phecode representations of our drug and disease data, respectively. Choosing these representations allowed us to balance data granularity and utility in optimizing model performance, as we tested associations of agent names (but not formulations, as would be available from ATC-5 embeddings) against high-level phenotype codes with logical mappings to the ICD ontology. While drug formulation could present interesting relationships to toxicity (e.g., through elevated concentrations at sensitive physiological sites like the cervix), we consider that our inability to capture this information does not detract from the power of our model to robustly capture associations between maternal drug exposures and adverse neonatal outcomes, as the explicit goal of our model was to discover relationships between the agents mothers consume and adverse outcomes in their neonates. In this way, we consider formulation to have trace effects on fetal toxicity, further given that most agents within our list of agents are consumed orally.

We observed replication of 8 well-known teratogens [phenytoin⁴⁹, valproate⁵⁰, fenofibrate⁵¹, quinapril⁵², retinoids (tazarotene, vitamin A, and adapalene)⁵³, and, topiramate⁵⁴] and 2 teratogens confirmed by our clinical consults [salicylates (phenyl salicylate and salicylic acid)^{55,56}] within our MedWAS results. We consider the according 22 signals (sample signal presented within Table 1, below) as positive population controls for our model: when we identified maternal medication history across our health system, we anticipated that such “anchoring” drugs would present with associations to neuroteratogenic outcomes. Negative population associations (i.e., prescription drugs with known protective effects against teratogenicity and/or zero baseline risk of teratogenic outcomes) are inherently uncommon and were therefore difficult for us to develop, further given that protective agents like folate are often taken by all expectant mothers receiving medical care during pregnancy, in addition to other potentially toxic drugs⁵⁷. Herein, our replication of positive control signals through MedWAS allowed us sufficient confidence to proceed with analysis of our model outcomes; our intention to develop structured statistical models with inherent controlling—both for each drug-disease test and across our population—also affirms our non-exploratory study design.

We considered Bonferroni significance a soft constraint, given increasing consensus that purely statistical significance does not directly imply biological significance—especially in the context of holistic approaches like PheWAS⁵⁸. Instead, we maintained signals with significant *p*-values at a baseline of 95% confidence even if they did not demonstrate Bonferroni significance, relying on the other evidentiary filters we describe below to determine their relative importance. This approach to determining signal significance holds in all places in which we discuss “significant” outcomes within this manuscript.

Top-Ranking Signals. With a list of convincing drug-disease hypotheses, anchored in statistical significance, literature evidence of preclinical and clinical toxicity, the norms of pediatric and obstetric practice, and replicative case series, we identified several classes of drugs with convincing signals of fetal toxicity that we believe warrant further assessment through more structured

epidemiological investigations. These demonstration signals demonstrate the utility of our MedWAS approach to generate a pliable, hypothesis-generating pipeline for the stimulation of post-market regulatory review programs for drug safety in pregnancy.

The following classes of drugs appeared most significantly linked to clusters of adverse neurological Phecodes diagnosable in the perinatal period, including “spina bifida” ($n = 219$ children), “neural tube defects” ($n = 242$ children), epilepsy and convulsions ($n = 2343$ children), abnormal (involuntary) movements ($n = 602$ children), (obstructive) sleep apnea ($n = 1,376$ children), and “infantile cerebral palsy” ($n = 149$ children). We present these relationships not solely from statistical results, but from considering the holistic evidence review that we describe above: (1) With limiting $p = 4 \times 10^{-10}$ and OR = 1.03, anti-epileptic drugs (including gabapentin, a drug routinely used off-label⁵⁹, and known toxicants like valproate and topiramate⁶⁰, as described above); (2) with limiting $p = 2 \times 10^{-7}$ and OR = 1.06, psychotropic agents (including alprazolam and other anxiolytic agents, which are often consumed by pregnant patients but have conflicting safety data on their labels⁶¹); (3) with limiting $p = 1 \times 10^{-4}$ and OR = 1.02, anti-emetic drugs (including ondansetron, which, while numerous studied in relationship to fetal cardiovascular outcomes⁶², is often consumed in the first trimester and features controversial associations to pediatric central nervous system abnormalities^{63,64}); (4) with limiting $p = 8 \times 10^{-8}$ and OR = 1.50, narcotic analgesics⁶⁵ (including fentanyl, which featured >60% coincidence rate between maternal drug exposure and detrimental neonatal phenotype and occurred with similar disease links and coincidence rates to the opiate antagonist naloxone); (5) with limiting $p = 4 \times 10^{-3}$ and OR = 1.83, anti-cancer drugs (including tamoxifen, a drug with few uses among pregnant people who choose not to terminate their pregnancies upon a cancer diagnosis—despite its narrow therapeutic index, the drug does not feature a contraindication for pregnancy on its label⁶⁶).

Our teratology quantitative structure-activity relationship model that we describe in “Introduction”²⁶ concurred with our present analysis of drugs containing fluoroquinolone and azetidinone motifs, providing us with an additional layer of validative evidence review in support of the performance of our process.

Discussion

Our results demonstrate that systematic assessment of the pharmacological determinants of pregnancy outcomes is possible via RWE synthesis that repurposes information routinely collected from primary care and is sufficiently flexible to accommodate direct input from the clinical stakeholders who provide care to pregnant patients and their newborn children. In this regard, our process presents the importance of complementing quantitative methods with qualitative evidence, as much of the contextual knowledge on obstetric prescriptive practice and pediatric disease assessment remains unavailable in structured databases. This combination of ML and consensus prioritization among human users for accurate outcomes analysis is archetypal of PheWAS

and GWAS approaches, as many previous publications affirm^{24,38,67}.

Our signals present opportunities for confirmation and further interrogation through mechanistic models of human development, as well as for more rigorous evaluation through regulatory-facing program development^{68–70}. This expansion is facilitated by the availability of an ontology of medical record numbers for patients with each drug exposure and each outcome that we tested, facilitating a review of individual EHRs to confirm true incidence of prescription and disease, as well as to understand confounding variables within the natural history of patients' care that our quality control system did not consider but may otherwise explain disease signals. These chart reviews are important and must be undertaken rigorously (e.g., through a repeated random sampling approach) for each drug class in which there is interest in the deeper study. In this way, continuously integrating knowledge about the clinical implementation of implicated agents and the manifestations of their related diseases will allow for further specification of our hypothesis generation platform in the more probative research that we have planned in the future. Nonetheless, reliance on medication history in the EHR carries the risk of exposure misclassification, as some pregnant people may not consume medications as prescribed or may self-report their medical history incompletely or inaccurately. In this study, we attempted to mitigate the misclassification risk by parsing signals associated to OTC drugs and nutraceuticals, but we note that the above is an ontological limitation of the EHR. Sources of electronic drug exposure data reflecting increased patient interaction (e.g., therapeutic dose monitoring (TDM) systems) may only be available for certain drugs—and only contain data at specific timepoints—substantially increasing the likelihood of bias from data missingness that is resultant from the limited use of TDM in routine clinical practice.

We again affirm that the goal of our research was the development of an enterprise-wide, hypothesis-generating pipeline of drug safety signals, to inform post-regulatory authorization safety studies. This work does not aim to identify malpractice and does not comprise clinical guidance on prescriptive behavior for pregnant patients.

Despite this orientation and the advantages of our approach, our methods have important limitations that can also spark new research questions. Beyond the randomization barriers we describe in “Introduction,” ontological barriers prevented us from executing PS to explicitly balance our cohorts before attempting MedWAS for the drug-disease inference within closely matched sub-groups. We considered alignment of maternal morbidity to the Charlson comorbidity index⁷¹ and application of the superficial method of PS developed by Choi et al. for PheWAS-empowered drug development studies⁷², to match mothers with similar baseline medical and demographic histories for comparison through MedWAS. While, if successful, this approach could have increased the resiliency of our analyses to confounding from variables extraneous to the prescription of the drug specified for each experiment, we realized that the number of maternal-fetal linkages from a single academic medical center like ours is too low to achieve the maximal level of controlling *in situ*. While ~100 K EHRs is a moderately-large data set for implementation of the present research—and represents the data captured from a large, productive medical center—this project demonstrated that execution of our methods with automated controls for confounding by maternal disease history and patient demography requires access to larger databases to prevent attrition of all comparable patient records. Though we could not employ PS *in situ*, as we had originally hoped, we believe that the evidence synthesis workflow we developed—along with the availability of manual patient chart review modules alongside MedWAS—

successfully helped us to address the effects of these potentially confounding variables through our signal vetting and prioritization procedure. In future research, we hope to access larger administrative databases of patient records, so we may better integrate PS into our quantitative process. This access could also allow facilitate testing against more discrete representations of neonatal phenotypes than those encoded by Phecodes.

We affirm throughout this manuscript that a central challenge to studying pregnancy and its outcomes with EHRs is defining the period of gestation. Many EHR systems rely on a “pregnancy flag,” encoding, on the backend, a binary representation of pregnancy status⁷³. This flag is problematic^{74,75}, as we have noticed in our EHR system that it often triggers by elevation in a patient's body mass index. Therein, reduced precision from the available marker means that inferential approaches to defining the period of pregnancy are necessary to layer other study elements, such as identifying a patient's medication history during gestation. Arithmetic approaches—such as subtracting 40 weeks from a patient's delivery date documented on a labor and delivery form to estimate conception date—are possible for first-pass estimation of gestational period, but they rely on low missingness in delivery date information within a candidate EHR data set. Contrastingly, as we describe in “Methods,” we found that an inferential approach to predicting the first date of gestation is a plausible pathway for pregnancy identification, as data missingness in the extraction of delivery date from the provider-facing EHR to institutionally maintained databases for secondary use is surprisingly significant. Our approach is also more accurate than the arithmetic alternative we describe above, as the former relies on multiple validated signals of obstetric care. We consider this approach more parsimonious than one of systematic data imputation followed by arithmetic determination, and we affirm that detailed informatics of pregnancy determination in the EHR lie outside the scope of the present study (here, we sought a minimum viable solution that could facilitate MedWAS). Similarly, we are unaware of a row-level data source on the natural history of pregnancy that does not present such ontological limitations or that does not require statistical approaches to defining gestational time. Our approach (including gestational period definition and MedWAS execution) is sufficiently robust to work across other data sets aligned to the Observational Medical Outcomes Partnership common data model (CDM)⁷⁶, while enabling our analysts to readily reproduce our phenotyping for future experiments at our institution, given the approach's training on our EHR data. Nonetheless, refinement of our pregnancy identification approach (to improve its accuracy and robustness for more complex test cases) is an area for future development, which would benefit from collaboration with experts on data standards.

The context of our approach is most immediately aligned with discovering teratogenic associations, with more limited applicability to evaluating potential determinants of general patient safety outside those we considered in this work; this argument is based, in part, on our model's reliance on teratogens as positive data controls. Also, by design, our study evaluates perinatal outcomes, as testing relationships between *in utero* drug exposures and phenotypes at prolonged stages of the pediatric life course remains highly difficult due to the accumulation of potentially confounding etiologies during the natural history of childhood⁷⁷. Gaps between the informatics strategy underlying our model and the clinical context of prescriptive practice during pregnancy could further restrict the utility of this approach—while some drugs significantly implicated by our model may create transient neurological disturbances in infants, prescriptive practice during pregnancy is most likely to consider long-term risk to the fetus against immediate therapeutic benefit to the

mother; this balance is difficult to evaluate with the ontological limitations of EHR data. For these reasons—alongside the primary goal of our work to study safety outcomes—this research does not seek to quantitatively discuss the efficacy vs. safety profiles (i.e., therapeutic indices) of maternal drug exposures.

While we do not consider the boundaries of our phenotyping capabilities as a significant limitation of our approach, we believe that quantifying the extent to which drug exposures during pregnancy can create lifelong disabilities is an important question. Addressing this question remains a “grand challenge” in the fields of pharmacoepidemiology and life course research and therefore warrants the creation of data management infrastructure that is more capable of reliably capturing patients’ childhood progressions through a collection of systems more diverse than EHRs⁷⁸. Nonetheless, we affirm that our decision to implement MedWAS across all pediatric outcomes, with downstream filtration of results to only perinatal outcomes, allowed us to accomplish our goal of discovering potentially iatrogenic etiologies for birth defects, while also allowing us to prospectively harness our data for studies of prenatal determinants of adverse health outcomes that manifest later in childhood, if we can access other data types that we could harmonize with our model’s results.

Similarly, we considered drug exposure during the entire gestational period to enable MedWAS, as we could generate an outcome set of signals associated to a diversity of neonatal diseases from one execution of the model. *Post hoc*, as we evaluated our signals for a neurological disease test case, we restricted our signals to those only associated to drugs with evidence of maternal use during the first trimester of pregnancy. The benefit of this staged approach is that for any future studies that interrogate neonatal diseases associated with pathophysiology that manifests during a different gestational period, we may re-visit one, holistic data set generated by our MedWAS and restrict per a different time window of maternal drug exposure, to hypothesize a list of potential pharmaceutical determinants of that outcome. Herein, we may focus on signal interrogation for new use cases, rather than re-constructing the signal generation phase of our work.

Our understanding of the potential applicability of MedWAS towards new drug development is two-fold. First, we consider that drug candidates within the same class as existing drugs (i.e., chemical structure or biological indication) could feature similar safety profiles, as supported by the teratogenicity QSAR model that our group previously published referenced above²⁶. Second, we believe that the development of new therapeutic uses (NTUs) by label expansion (i.e., evaluation of the efficacy of an approved drug for a new indication) could benefit from MedWAS results in the consideration of whether to include pregnant people in prospective experiments for safety evaluation, as well as in determining the applicability of an NTU toward diseases within the pregnant population through leveraging existing, post-marketing data about the drug in its original use⁷⁰. Nonetheless, even within drug classes we evaluated, individual drugs’ PK could vary⁷⁹; therefore, our process is not designed to accommodate safety signal detection before phase IV of drug development.

In keeping with the results of most PheWAS studies, we are aware that the signals generated from this platform are potentially controversial⁸⁰ and that despite our attempts to integrate multiple streams of clinical, statistical, biological, and archival evidence with manual EHR review, several of the hypotheses we generated may be explained by non-pharmacological factors. We believe, however, that the strength of our platform is in the identification of priority areas for post-market review of drug use during pregnancy that is anchored in RWE, that is sufficiently robust to accommodate the diversity of maternal drugs and perinatal diseases that naturally manifest in a large health system, and that is

sufficiently parsimonious to allow for process replication in other health enterprises. We consider that the limited preconditioning necessary for the execution of our approach makes it packagable, and that the qualitative aspects of our study design allow us to engage necessary clinical stakeholders for drug review more closely than further automated approaches might.

Future research may take forward our high-level identification of potentially unsafe drug classes, through more structured, epidemiological probes of exposure and outcome. Similarly, we believe the standardization of our quantitative process makes it sufficiently pliable to implement at other health systems with the same CDM underlying their EHRs, which we hope will further advance our understanding of the robustness of our hypothesis generation approach, when deployed across multiple sites’ data warehouses.

We envision that this work will allow us to partner with regulators of drug products to develop new programs that harmonize real-world data sources, towards detecting and evaluating safety signals for drugs authorized for use among pregnant patients. Furthermore, as part of a bench-to-bedside initiative to generate more accurate signals of drug safety in the regulatory evaluation of drug products potentially prescribable to pregnant people, we are currently developing organotypic models of the human placenta⁸¹ and developing brain⁸² that can allow us to validate our most convincing MedWAS signals on a mechanistic basis.

Data availability

Disclosure of our MedWAS data, though de-identified and aggregated, is subject to approval and oversight by the Office of Contracts Management (OCM) at VUMC, as our source data is derived from protected health information (PHI), and some drug-disease pairs are individually re-identifiable. Therefore, institutional policies prevent us from publicly releasing our data tables and their annotations in the interest of patient data security, but, within the data sharing regulations of our institution, we have attempted to provide meaningful information on the content and formatting of our outputs throughout this manuscript. Linked to our Supplementary Information file, we have also provided a supplementary data attachment (Supplementary Data 1), which provides a listing of agent tokens resultant from the phenotyping we describe above, alongside the number of pregnant people containing each agent within their EHRs (agents with no more than 5 exposures are censored accordingly within the spreadsheet). We are committed to open-source science and to ensuring the reproducibility of the research we present here; therefore, we are happy to discuss data transfer requests with researchers interested in our results. Interested investigators should contact the Corresponding Authors at the addresses accompanying this manuscript, and they are happy to discuss forwarding such requests to OCM (towards a data use agreement) within 30 days of receiving such a collaboration request.

Code availability

We leveraged a suite of previously published natural language processing (NLP) tools to extract phenotypic attributes and maternal drug exposures from narrative EHR data; versions, accession links, and source paper citations for these resources (which contain further description of their mechanisms) are provided at the points of their first reference within the text of our manuscript. The PheWAS package in R, version 0.12.3, is open-source and available through the GitHub link provided above, as well as further described in the PheWAS source papers cited throughout this manuscript. The custom implementation of this open-source package for MedWAS is governed by the same disclosure policy as our MedWAS data; access to this code may be requested through the mechanism we describe in “Data Availability” (we describe the execution in R version 3.6.3, via cookbook format, within our accompanying Supplementary Methods).

Received: 14 November 2021; Accepted: 1 September 2022;

Published online: 16 September 2022

References

1. Jamkhande, P. G., Chintawar, K. D. & Chandak, P. G. Teratogenicity: a mechanism based short review on common teratogenic agents. *Asian Pacific J. Tropical Dis.* **4**, 421–432 (2014).
2. Wood, W. FDA pregnancy categories: help or hindrance? *Mental Health Clinician* **3**, 78–80 (2013).

3. Pregnancy and Lactation Labeling (Drugs) Final Rule. *FDA* (2021).
4. Garrison, J. A. UpToDate. *J. Med. Libr. Assoc.* **91**, 97 (2003).
5. Riley, L. E., Cahill, A. G., Beigi, R., Savich, R. & Saade, G. Improving safe and effective use of drugs in pregnancy and lactation: workshop summary. *Am J Perinatol* **34**, 826–832 (2017).
6. Ram, D., Gowdappa, B., Ashoka, H. G. & Eiman, N. Psychopharmacoteratophobia: excessive fear of malformation associated with prescribing psychotropic drugs during pregnancy: an Indian perspective. *Indian J Pharmacol.* **47**, 484–490 (2015).
7. Smith, D. D. et al. Exclusion of pregnant women from clinical trials during the coronavirus disease 2019 pandemic: a review of international registries. *Am J Perinatol* **37**, 792–799 (2020).
8. van der Graaf, R. et al. Fair inclusion of pregnant women in clinical trials: an integrated scientific and ethical approach. *Trials* **19**, 78 (2018).
9. Shields, K. E. & Lyerly, A. D. Exclusion of pregnant women from industry-sponsored clinical trials. *Obstet Gynecol* **122**, 1077–1081 (2013).
10. Grzeskowiak, L. E., Gilbert, A. L. & Morrison, J. L. Methodological challenges in using routinely collected health data to investigate long-term effects of medication use during pregnancy. *Ther. Adv. Drug Saf.* **4**, 27–37 (2013).
11. Challa, A. P. et al. EHRs could clarify drug safety in pregnant people. *Nat. Med.* **1–2**, <https://doi.org/10.1038/s41591-020-0925-1> (2020).
12. Pulley, J. M. et al. Using what we already have: uncovering new drug repurposing strategies in existing omics data. *Annu. Rev. Pharmacol. Toxicol.* <https://doi.org/10.1146/annurev-pharmtox-010919-023537> (2019).
13. Wilmer, E., Chai, S. & Kroupouzou, G. Drug safety: pregnancy rating classifications and controversies. *Clin. Dermatol.* **34**, 401–409 (2016).
14. Huybrechts, K. F., Bateman, B. T. & Hernández-Díaz, S. Use of real-world evidence from healthcare utilization data to evaluate drug safety during pregnancy. *Pharmacoepidemiol. Drug Saf.* **28**, 906–922 (2019).
15. Ross, M. E. et al. Propensity score methods for analyzing observational data like randomized experiments: challenges and solutions for rare outcomes and exposures. *Am. J. Epidemiol.* **181**, 989–995 (2015).
16. Casey, J. A., Schwartz, B. S., Stewart, W. F. & Adler, N. E. Using electronic health records for population health research: a review of methods and applications. *Annu. Rev. Public Health* **37**, 61–81 (2016).
17. Cheung, K. et al. The concordance between self-reported medication use and pharmacy records in pregnant women. *Pharmacoepidemiol. Drug Saf.* **26**, 1119–1125 (2017).
18. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
19. Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
20. Franklin, J. M. et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. *Circulation* **143**, 1002–1013 (2021).
21. Groenwold, R. H. H. Trial emulation and real-world evidence. *JAMA Network Open* **4**, e213845 (2021).
22. Caniglia, E. C. et al. Emulating a target trial of antiretroviral therapy regimens started before conception and risk of adverse birth outcomes. *AIDS* **32**, 113–120 (2018).
23. Platt, R. W. Invited commentary on the study of medications in pregnancy. *Curr. Epidemiol. Rep.* **3**, 265–267 (2016).
24. Challa, A. P. et al. Systematically prioritizing candidates in genome-based drug repurposing. *ASSAY Drug Dev. Technol.* **17**, 352–363 (2019).
25. Challa, A. P. et al. Systematically prioritizing targets in genome-based drug repurposing. In: *Proc. 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics—BCB '18* 543–543 (ACM Press, 2018). <https://doi.org/10.1145/3233547.3233651>.
26. Challa, A. P. et al. Machine learning on drug-specific data to predict small molecule teratogenicity. *Reproductive Toxicol.* **95**, 148–158 (2020).
27. PheWAS - Phenome Wide Association Studies. <https://phewascatalog.org/phecodes>.
28. Danciu, I. et al. Secondary use of clinical data: the vanderbilt approach. *J Biomed Inform* **52**, 28–35 (2014).
29. Ramirez, A. H. et al. Modulators of normal electrocardiographic intervals identified in a large electronic medical record. *Heart Rhythm.* **8**, 271–277 (2011).
30. Denny, J. C., Irani, P. R., Wehbe, F. H., Smithers, J. D. & Spickard, A. The KnowledgeMap Project: development of a concept-based medical school curriculum database. *AMIA Annu. Symp. Proc.* **2003**, 195–199 (2003).
31. Denny, J. C., Smithers, J. D., Miller, R. A. & Spickard, A. ‘Understanding’ medical school curriculum content using KnowledgeMap. *J. Am. Med. Inform. Assoc.* **10**, 351–362 (2003).
32. Doan, S., Bastarache, L., Klimkowski, S., Denny, J. C. & Xu, H. Integrating existing natural language processing tools for medication extraction from discharge summaries. *J. Am. Med. Inform. Assoc.* **17**, 528–531 (2010).
33. Denny, J. C. et al. Evaluation of a method to identify and categorize section headers in clinical documents. *J. Am. Med. Inform. Assoc.* **16**, 806–815 (2009).
34. Xu, H. et al. MedEx: a medication information extraction system for clinical narratives. *J. Am. Med. Inform. Assoc.* **17**, 19–24 (2010).
35. Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucl. Acids Res.* **32**, D267–D270 (2004).
36. Nelson, S. J., Zeng, K., Kilbourne, J., Powell, T. & Moore, R. Normalized names for clinical drugs: RxNorm at 6 years. *J. Am. Med. Inform. Assoc.* **18**, 441–448 (2011).
37. Denny, J. C. et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1111 (2013).
38. Jerome, R. N. et al. Using Human ‘Experiments of Nature’ to Predict Drug Safety Issues: an example with PCSK9 inhibitors. *Drug Saf.* **41**, 303–311 (2018).
39. Liao, K. P. et al. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non-rheumatoid arthritis controls. *Arthritis Rheum.* **65**, 571–581 (2013).
40. Safarova, M. S. et al. A phenome-wide association study to discover pleiotropic effects of PCSK9, APOB, and LDLR. *NPJ Genom Med.* **4**, 3 (2019).
41. Anatomical Therapeutic Chemical (ATC) Classification. <https://www.who.int/tools/atc-ddd-toolkit/atc-classification>.
42. Wei, W.-Q. et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS ONE* **12**, e0175508 (2017).
43. Wasserstein, R. L. & Lazar, N. A. The ASA statement on p-values: context, process, and purpose. *American Statistician* **70**, 129–133 (2016).
44. Monte, A. A. et al. The accuracy of electronic medical record medication reconciliation in emergency department patients. *J. Emerg. Med.* **49**, 78–84 (2015).
45. Reprotox • <https://reprotox.org/member>.
46. About Southern Crossroads. *Southern Crossroads SoX* <https://www.sox.net/about-us/>.
47. van Gelder, M. M. H. J. et al. Teratogenic mechanisms of medical drugs. *Hum. Reprod. Update* **16**, 378–394 (2010).
48. Bhandari, J. & Thada, P. K. In *StatPearls* (StatPearls Publishing, 2021).
49. Fetal Hydantoin Syndrome. NORD (National Organization for Rare Disorders) <https://rarediseases.org/rare-diseases/fetal-hydantoin-syndrome/>.
50. Fetal Valproate Syndrome. NORD (National Organization for Rare Disorders) <https://rarediseases.org/rare-diseases/fetal-valproate-syndrome/>.
51. Familial Hypercholesterolemia and Pregnancy. American College of Cardiology <https://www.acc.org/latest-in-cardiology/articles/2018/05/10/13/51/http%3a%2f%2fwww.acc.org%2flatest-in-cardiology%2farticles%2f2018%2f05%2f10%2f13%2f51%2ffamilial-hypercholesterolemia-and-pregnancy>.
52. ACCUPRIL®(quinapril HCl) | Pfizer Medical Information - US. <https://www.pfizermedicalinformation.com/en-us/accupril>.
53. Updated measures for pregnancy prevention during retinoid use. 4.
54. FDA Center for Drug Evaluation and Research. FDA Drug Safety Communication: Risk of oral clefts in children born to mothers taking Topamax (topiramate). (2019).
55. Gould, W. Pregnant? Stop Using This Common Skincare Ingredient ASAP. *POPSUGAR Beauty* <https://www.popsugar.com/node/44510196> (2018).
56. Beauty Ingredients to Avoid During Pregnancy | Vogue. <https://www.vogue.com/article/beauty-ingredients-pregnancy-safety-toxic>.
57. Greenberg, J. A., Bell, S. J., Guan, Y. & Yu, Y. Folic acid supplementation and pregnancy: more than just neural tube defect prevention. *Rev. Obstet. Gynecol.* **4**, 52–59 (2011).
58. Challa, A. P. et al. Human and machine intelligence together drive drug repurposing in rare diseases. *Front. Genet.* **12**, 1374 (2021).
59. Peckham, A. M., Evoy, K. E., Ochs, L. & Covvey, J. R. Gabapentin for off-label use: evidence-based or cause for concern? *Subst. Abuse* **12**, 1178221818801311 (2018).
60. Hernández-Díaz, S. et al. Comparative safety of antiepileptic drugs during pregnancy. *Neurology* **78**, 1692–1699 (2012).
61. Alprazolam: Drug information - UpToDate. https://www.uptodate.com/contents/alprazolam-drug-information?search=alprazolam&source=panel_search_result&selectedTitle=1~121&usage_type=panel&kp_tab=drug_general&display_rank=1.
62. Lavecchia, M., Chari, R., Campbell, S. & Ross, S. Ondansetron in pregnancy and the risk of congenital malformations: a systematic review. *J. Obstet. Gynaecol. Canada* **40**, 910–918 (2018).
63. Huybrechts, K. F. et al. Association of maternal first trimester ondansetron use with cardiac malformations and oral clefts in offspring. *JAMA* **320**, 2429–2437 (2018).
64. Brain Defect Lawsuit | Zofran Linked To Major Birth Defects. <http://zofran.monheit.com/zofran-brain-defects/>.

65. CDC. Basics About Opioid Use During Pregnancy|CDC. Centers for Disease Control and Prevention <https://www.cdc.gov/pregnancy/opioids/basics.html> (2021).
66. Therapeutic choice for Tamoxifen patients|Soltamox (tamoxifen citrate). Soltamox® <https://soltamox.com/offering-tamoxifen-patients-a-therapeutic-choice/>.
67. Pulley, J. M. et al. When enough is enough: decision criteria for moving a known drug into clinical testing for a new indication in the absence of preclinical efficacy data. *Assay Drug Dev. Technol.* **15**, 354–361 (2017).
68. EHRs + Machine Learning Decipher Drug Effects In Pregnant Persons. Pubs-Clinical Research News Online. <https://www.clinicalresearchnews.com/news/2020/06/22/ehrs-machine-learning-decipher-drug-effects-in-pregnant-persons>.
69. Govern, P. Target trials support drug safety in pregnant patients. Vanderbilt University. <https://news.vumc.org/2020/05/27/target-trials-support-drug-safety-in-pregnant-patients/>.
70. Pre-term deliveries due to COVID-19 could be avoided by studying EHRs. EurekAlert! https://www.eurekalert.org/pub_releases/2020-06/vumc-dd061020.php.
71. Austin, S. R., Wong, Y.-N., Uzzo, R. G., Beck, J. R. & Egleston, B. L. Why summary comorbidity measures such as the Charlson Comorbidity Index and Elixhauser score work. *Med Care* **53**, e65–e72 (2015).
72. Choi, L. et al. Evaluating statistical approaches to leverage large clinical datasets for uncovering therapeutic and adverse medication effects. *Bioinformatics* **34**, 2988–2996 (2018).
73. Pregnancy status. LOINC <https://loinc.org/loinc/82810-3/>.
74. Representing Patient Pregnancy Status|Interoperability Standards Advisory (ISA). <https://www.healthit.gov/isa/representing-patient-pregnancy-status>.
75. Liyanage, H., Williams, J., Byford, R. & Lusignan, S. Ontology to identify pregnant women in electronic health records: primary care sentinel network database study. *BMJ Health Care Inform.* **26**, e100013 (2019).
76. OMOP Common Data Model—OHDSI. <https://www.ohdsi.org/data-standardization/the-common-data-model/>.
77. Nicolau, B. & Marcenes, W. How will a life course framework be used to tackle wider social determinants of health? *Community Dent. Oral Epidemiol.* **40**, 33–38 (2012).
78. Jones, N. L. et al. Life course approaches to the causes of health disparities. *Am. J. Public Health* **109**, S48–S55 (2019).
79. Tricco, A. C. et al. Comparative safety of serotonin (5-HT₃) receptor antagonists in patients undergoing surgery: a systematic review and network meta-analysis. *BMC Med.* **13**, 142 (2015).
80. Challa, A. P. et al. Human and machine intelligence together drive drug repurposing in rare diseases. *Front. Genet.* **0**, (2021).
81. Richardson, L. et al. Fetal membrane organ-on-chip: an innovative approach to study cellular interactions. *Reprod. Sci.* 1933719119828084, <https://doi.org/10.1177/1933719119828084> (2019).
82. Bosworth, A. M., Faley, S. L., Bellan, L. M. & Lippmann, E. S. Modeling neurovascular disorders and therapeutic outcomes with human-induced pluripotent stem cells. *Front. Bioeng. Biotechnol.* **5**, 87 (2018).

Acknowledgements

A.P.C., D.M.A., and E.A.G. acknowledge support from award R21HD105304 from the National Institute of Child Health and Development (NICHD) of NIH. A.P.C., D.M.A., and E.S.L. also acknowledge relevant funding from award U54TR02243-02 from the National Center for Advancing Translational Sciences (NCATS) of NIH. A.P.C. and D.M.A. were additionally funded for a study of therapeutic index in pregnancy by award N009367701 from Rainwater Charitable Foundation. S.L.V. was supported by IRSA

#1015006 from the Burroughs Wellcome Fund. The authors thank Noel Southall, PhD, Team Leader in NCATS's Division of Preclinical Innovation and the NCATS-FDA interagency initiative, for his advice on the framing and applications of the work we describe in this manuscript. We also thank Andrew Beam, PhD at Harvard T.H. Chan School of Public Health for his feedback on study design and results interpretation and Meghan Vance at VUMC for her help with developing Fig. 1. We acknowledge that the perspectives we present in this manuscript are solely the authors'; they do not necessarily represent the official views of NCATS, NICHD, or any other organization.

Author contributions

A.P.C. led the study team. A.P.C., X.N., L.M.B., E.S.L., R.R.L., and J.A.G. designed the platform described in this manuscript. X.N. and L.M.B., as the data analysts for this project, determined gestational ages, executed MedWAS and its associated sensitivity analyses, and evaluated the compatibility of PS matching with our data source. A.P.C., E.A.G., E.S.L., R.R.L., and J.A.G. developed the evidence synthesis procedure for MedWAS validation and signal prioritization. E.A.G., S.L.V., and D.M.A. provided clinical consultations on MedWAS outcomes. E.S.L. provided neuroscientific consultation on MedWAS outcomes, and R.R.L. provided pharmacological consultation. A.P.C. drafted this manuscript for the study team's review, and X.N. drafted the Supplementary Methods section that accompanies the manuscript. A.P.C. created Fig. 1, Supplementary Data 1, and the Supplementary Discussion accompanying this manuscript. All co-authors had an opportunity to review the manuscript and its supplement.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-022-00181-w>.

Correspondence and requests for materials should be addressed to Anup P. Challa or David M. Aronoff.

Peer review information *Communications Medicine* thanks Oliver Daly and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022