

# The Next Generation of Drug Safety Science: Coupling Detection, Corroboration, and Validation to Discover Novel Drug Effects and Drug–Drug Interactions

Nicholas P. Tatonetti<sup>1,2,3,4</sup>

**Rare adverse drug reactions and drug–drug interactions (DDIs) are difficult to detect in randomized trials and impossible to prove using observational studies. We must ascribe to a new way of conducting drug safety research that has the efficiency of a retrospective analysis and the rigor of a prospective trial. This can be achieved by integrating observational data from humans with laboratory experiments in model systems. The former establishes clinical significance and the latter supports causality.**

## BACKGROUND

Adverse drug reactions are a leading cause of morbidity and mortality around the world. The occurrence of severe and unexpected side effects reduces the usefulness of pharmaceutical treatments. Some reactions are common across all who are treated, but often adverse reactions only occur in a subpopulation of patients. DDIs are an example of this. The patients taking either drug alone will not experience the reaction, while the subset of patients exposed to both will. Pharmacogenetic variants that modulate drug response or comorbidities that increase the risk of adverse reactions are additional examples. The challenge, of course, is discovering these subpopulation risk factors without any prior knowledge of them.

Detection of subpopulation-specific adverse reactions is statistically challenging. Clinical trials do not have the power to study DDIs or to evaluate specific subgroups. On the other hand, while

retrospective studies of real-world clinical data have the diversity and the size to detect these patterns, they also have unavoidable biases and systematic errors<sup>1</sup> that produce dubious discoveries (**Box 1**). A new approach to drug safety research that takes the best aspects of both methods is needed—combining the validity of a prospective trial with the opportunity for discovery provided by observational data.

I propose the use of a three-step methodology (**Figure 1**) to discover and validate novel adverse drug reactions and DDIs. The first step, *detection*, is to data-mine a large observational resource, like the US Food and Drug Administration's (FDA) Adverse Event Reporting System (FAERS), for unexpected associations between drugs and adverse events. This procedure will produce thousands of significant hypotheses and, as is true in any data-mining experiment, a large proportion of those are expected to be false discoveries. There are

many signal detection and statistical data-mining algorithms to choose from when detecting new drug–adverse event hypotheses.<sup>2</sup> The precise choice of method will depend on the adverse event outcome of interest and the balance desired between false positives and novel discoveries.

The second step, *corroboration*, is to integrate an independent resource (e.g., chemical network biology data or electronic health records, EHRs) to prioritize the data-mined hypotheses based on their plausibility. A drug that can be molecularly connected to the adverse event is more likely to be true than one that cannot. Chemical informatics methods and systems pharmacology models that use drug binding and pathway data can be used to find these molecular connections.<sup>3,4</sup> Likewise, a drug effect that can be independently replicated in EHRs is more likely than one that cannot. Since this step is also performed on already collected data using computational methods, it can be applied to all of the mined hypotheses from the first step. Only those hypotheses that pass collaboration are kept. This will reduce the number of significant hypotheses from thousands to hundreds or tens. The limitation here is that these models are specific to the adverse reaction being studied. For example, the same molecular model—or EHR phenotype—used for studying arrhythmias will likely not apply to glucose metabolism.

In the final step, *validation*, a model system is identified to experimentally test the strongest hypotheses. This is the most challenging step, as many adverse reactions do not map clearly to experimental systems. It is also the most important, since a hypothesis is only as good as it is falsifiable. Ideally, the experiment will be efficient, with a straightforward interpretation, like a protein affinity assay. Often, however, a more complex experiment in a cellular or animal model is needed. In either case, this is both an efficient and ethical method for validating adverse drug reactions compared to launching a prospective clinical trial. These three sources of evidence, from detection to corroboration to validation, provide a compelling case for the newly discovered effect.

<sup>1</sup>Department of Biomedical Informatics, Columbia University, New York, New York, USA; <sup>2</sup>Department of Systems Biology, Columbia University, New York, New York, USA; <sup>3</sup>Institute for Genomic Medicine, Columbia University, New York, New York, USA; <sup>4</sup>Data Science Institute, Columbia University, New York, New York, USA. Correspondence: N.P. Tatonetti (nick.tatonetti@columbia.edu)

doi:10.1002/cpt.949

## Box 1.

Data-mining in a forest of bias. A simple thought experiment illustrates the difficulty introduced by biases in observational data. Consider a small forest populated by red birds and blue birds. Suppose that we would like to answer the question of which bird is more prevalent in this small forest. We decide to venture out and count what we see. Upon analysis of our data, we find that we observed five times as many blue birds as red birds. It is tempting to conclude that there are more blue birds than red birds in our hypothetical forest; however, this would be incorrect. In fact, we cannot make any reliable conclusions from these data. To explain, let me now reveal something about these birds. Blue birds are friendly and curious birds. When they hear people enter the forest they fly to them to investigate and show off. Red birds, on the other hand, are scared and grumpy birds. When *they* hear people enter the forest they fly away or hide. The behaviors of these birds, therefore, changes—or *confounds*—what we can observe, making the data we collected useless for the question we wanted to answer. This is not a hypothetical problem. This is exactly what occurs every day in every clinic and hospital when patient data are collected in the EHR. What we can observe is dictated by the behaviors of the patients, doctors, and nurses. Patients do not admit themselves to the hospital at random and neither do physicians prescribe medications that way.

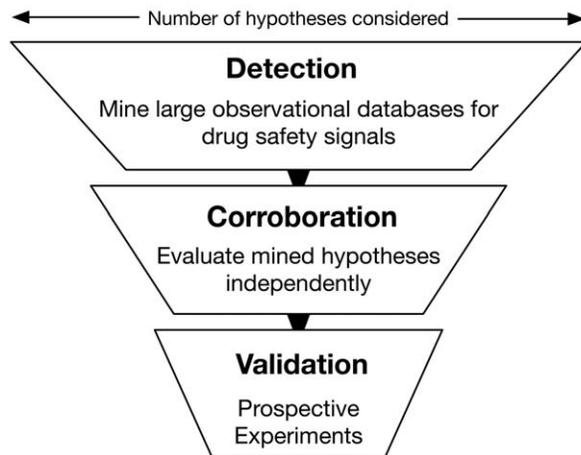
## DISCOVERY IN PRACTICE

This strategy of detection, corroboration, and validation is not hypothetical, but one that my colleagues and I have used successfully. First, to show that comedication of paroxetine and pravastatin leads to increased blood glucose,<sup>5</sup> and later to prove that the combination of ceftriaxone and lansoprazole prolongs the QT interval.<sup>6</sup> In both cases, we used Latent Signal Detection<sup>7</sup> to mine unexpected DDIs from FAERS. Latent Signal Detection is a supervised machine-learning algorithm that can identify associations with adverse reactions even if they are not explicitly reported. We trained one model to identify glucose dysregulation<sup>5</sup> and the other for acquired long QT syndrome (LQTS).<sup>7</sup> Hundreds of putative DDIs were hypothesized for each effect. In our experience, as many as 90–95% of these hypotheses will be false positives, underscoring the importance of a computational method to corroborate these putative drug effects for plausibility.

To corroborate, or refute, the data-mined hypotheses we used an independent analysis of EHR data. For glucose dysregulation (GD), we pulled glucose lab results run before and after exposure to the drugs, individually and in combination. Likewise, for LQTS we extracted the corrected QT value from the electrocardiogram reports of patients exposed to either or both drugs. In both, the analysis we used was efficient and could be applied to each of the mined hypotheses relatively quickly, enabling us to eliminate most of the hypotheses as implausible. At the end of this stage, we have corroborated DDIs with evidence of a clinical effect in humans from two independent analyses based on two different observational datasets. Of those that remained, we prioritized them by their effect sizes and potential for clinical impact for experimental validation.

The final step, validation, is used to prove that these associations are due to the combination of drug exposures and not because of some confounding variable. We attempted validation on the top prediction from each model (paroxetine and pravastatin for GD and ceftriaxone and lansoprazole for LQTS). In neither case did we know the mechanism of the interaction and we took two different strategies for validation. For paroxetine and pravastatin, we

## Three steps for better drug safety surveillance



**Figure 1** The three steps to more rigorous and efficient drugs safety surveillance. Step 1: Detection. Modern signal detection and statistical data-mining methods are used to identify new drug safety and drug–drug interaction signals. Methods include traditional approaches, like disproportionality analysis with statistical corrections, and newer methods, like supervised machine-learning and pattern detection. This step produces a lot of statistically significant associations but also a lot of false discoveries. Step 2: Corroboration. Mined hypotheses are evaluated against an independent dataset and evaluated for plausibility. These additional data could score drug effect hypotheses by their molecular connection to the known effect or provide additional clinical evidence from alternative resources. Any hypotheses that does not corroborate is removed from consideration, greatly reducing the number of false discoveries. At the end of this stage only a few to dozens of the strongest hypotheses remain. Step 3: Validation. Corroborated hypotheses are validated using a model system. Model systems may be molecular assays (e.g., chemical-protein binding affinity), cellular systems, or animal models, depending on which model is best suited for the predicted adverse reaction outcome. Drug effects that validate in all three steps have demonstrated clinical importance in humans (Steps 1 and 2) and have evidence of causality (Step 3). This process is both more efficient than a clinical trial and more rigorous than a retrospective analysis.

used an insulin-resistant mouse model, giving us the greatest chance to observe the hypothesized effect regardless of if the interaction occurred at metabolism, glucose transport, insulin secretion, or gluconeogenesis. For ceftriaxone and lansoprazole, we chose a cellular assay of the drugs' effect on the hERG channel. This potassium ion channel is essential in heart rhythm and the blocking of this channel is the mechanism of action for all drugs known to prolong the QT interval. In both cases, the DDI validated<sup>5,6</sup> and in both cases it was the only predicted DDI we tested—resource limitations prohibited the evaluation of additional hypotheses.

The three steps of detection, corroboration, and validation produce findings that are reproducible. Our discovery that comedication of paroxetine and pravastatin increases blood glucose has been subsequently reproduced in rats<sup>8</sup> and humans<sup>9</sup> by independent groups. Similarly, in response to our publication of ceftriaxone and lansoprazole prolonging the QT interval, additional case reports have surfaced in support of the interaction.<sup>10</sup>

## DISCUSSION

The drug safety community is facing an insurmountable challenge when trying to accurately identify rare adverse reactions and drug combination effects using current pharmacovigilance methods. There are simply too many covariates, strata, and interactions for these effects to be discovered in the clinical trial phase. Retrospective methods are sensitive to biases and produce far too many false positives. The three-step process of detection and corroboration in human data coupled with validation in experimental model systems is a balance of

the two—more efficient than randomized controlled trials and more rigorous than observational analysis. The strategy of integrating computational analysis with experiments is not unique to drug safety science. Others have used similar approaches in drug repositioning, drug target prediction, and drug design.

Ultimately, the goal of any pharmacovigilance study is to alter clinical practice to improve patient safety. This is often achieved through actions taken by regulatory agencies like the FDA and the European Medicines Agency. These agencies would, of course, prefer evidence sourced from randomized controlled trials when issuing warnings or requiring withdrawals. In many cases, however, the data are not available. Discoveries supported by the three levels of evidence I present here should be considered adequate to make product safety decisions. The human data from spontaneous reporting systems and EHRs establishes the clinical importance of the effect (or the lack thereof, if that is what the data show). The prospective experiment in a model system establishes that the clinical effect observed is being caused by the hypothesized agent. This combination of clinical significance and causality evidence is as strong as a well-powered randomized trial and represents the most viable path forward for meaningful and impactful drug safety science.

## ACKNOWLEDGMENT

The author is funded by NIGMS R01 GM107145 and NCATS OT3 TR002027.

## CONFLICT OF INTEREST

The author is a paid advisor to Advera Health, Inc. The author declared no other conflict of interest.

© 2017 ASCPT

1. Hripcsak, G. & Albers, D.J. Next-generation phenotyping of electronic health records. *J. Am. Med. Inform. Assoc.* **20**, 117–121 (2013).
2. van Puijnenbroek, E.P. *et al.* A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol. Drug Saf.* **11**, 3–10 (2002).
3. Lorberbaum, T. *et al.* Systems pharmacology augments drug safety surveillance. *Clin. Pharmacol. Ther.* **97**, 151–158 (2014).
4. Jacunski, A. & Tatonetti, N.P. Connecting the dots: Applications of network medicine in pharmacology and disease. *Clin. Pharmacol. Ther.* **94**, 659–669 (2013).
5. Tatonetti, N.P. *et al.* Detecting drug interactions from adverse-event reports: Interaction between paroxetine and pravastatin increases blood glucose levels. *Clin. Pharmacol. Ther.* **90**, 133–142 (2011).
6. Lorberbaum, T. *et al.* Coupling data mining and laboratory experiments to discover drug interactions causing QT prolongation. *J. Am. Coll. Cardiol.* **68**, 1756–1764 (2016).
7. Lorberbaum, T., Sampson, K.J., Woosley, R.L., Kass, R.S. & Tatonetti, N.P. An integrative data science pipeline to identify novel drug interactions that prolong the QT interval. *Drug Saf.* **39**, 433–441 (2016).
8. Li, F. *et al.* Co-administration of paroxetine and pravastatin causes deregulation of glucose homeostasis in diabetic rats via enhanced paroxetine exposure. *Acta Pharmacol. Sin.* **35**, 792–805 (2014).
9. An, L., Ravindran, P.P., Renukunta, S. & Denduluri, S. Co-medication of pravastatin and paroxetine—a categorical study. *J. Clin. Pharmacol.* **53**, 1212–1219 (2013).
10. Lazzarini, P.E., Bertolozzi, I. & Rossi, M. Combination Therapy With Ceftriaxone and Lansoprazole, Acquired Long QT Syndrome, and Torsades de Pointes Risk. <[http://www.onlinejacc.org/content/69/14/1876.full?sso=1&sso\\_redirect\\_count=5&access\\_token=>](http://www.onlinejacc.org/content/69/14/1876.full?sso=1&sso_redirect_count=5&access_token=>)>.