nSides: An interactive drug–side effect gateway

Rami S. Vanguri
Department of Biomedical Informatics
Columbia University
New York, New York 10032
Email: r.vanguri@columbia.edu

Joseph D. Romano
Department of Biomedical Informatics
Columbia University
New York, New York 10032
Email: jdr2160@cumc.columbia.edu

Tal Lorberbaum
Department of Biomedical Informatics
Columbia University
New York, New York 10032
Email: tl2542@cumc.columbia.edu

Choonhan Youn
San Diego Supercomputer Center
University of California, San Diego
La Jolla, California 92092
Email: cyoun@sdsc.edu

Victor Nwankwo
Department of Biomedical Informatics
Columbia University
New York, New York 10032
Email: vtn2106@cumc.columbia.edu

Nicholas P. Tatonetti
Department of Biomedical Informatics
Columbia University
New York, New York 10032
Email: nick.tatonetti@columbia.edu

***Abstract*—Adverse drug events are a leading cause of morbidity and mortality around the world. Regulatory agencies, such as the Food and Drug Administration (FDA), maintain large collections of adverse event reports. These spontaneous reporting systems provide an opportunity to retrospectively study drug effects from patient population data. To make our results more accessible, we present a gateway to a new database of drug interactions (nSides) developed using the FDA Adverse Event Reporting System. Side effects for a variety of drugs and drug interactions are presented. Additionally, for high order drug combinations, with the support of the Science Gateway Community Institute's Extended Developer Support, we develop a novel system to request new models to be generated using the Open Science Grid on-demand.**

I. INTRODUCTION

Spontaneous reporting systems such as the FDA Adverse Event Reporting System (FAERS) are important resources for detecting drug adverse events after a drug is approved (pharmacovigilance). However, pharmacovigilance algorithms often lead to many false positive and false negative findings due to issues of confounding, and detection of drug-drug interactions is an even greater challenge. We previously developed databases for off-label drug effects (OFFSIDES) and drug interactions (TWOSIDES) that account for these limitations using the novel Statistical Correction for Uncharacterized Bias (SCRUB) [2] algorithm (described below, in §II), both of which are publicly available for download. Here, we present a web gateway which acts an interface to a database of drug effects calculated using the previously developed SCRUB algorithm, known as nSides (https://nsides.io). nSides aims to make the database of drug effects accessible to researchers, clinicians, and patients, and contains features related to drug safety. Since it is not feasible to generate models for every possible drug combination, we develop a novel model request system which submits jobs to the Open Science Grid and appends the results to the databases for future access.

Another popular resource for drug effects is SIDER [6], which is generated using data from placebo-controlled clinical trials, biomedical literature, and electronic health records. Since the data sources for SIDER are separate from FAERS, which is used generate results for nSides, SIDER is used to validate the SCRUB algorithm.

II. DATA SOURCES AND METHODOLOGY

nSides makes use of several data sources. We use a curated version of the FDA Adverse Event Reporting System (FAERS) known as Adverse Event Open Learning through Universal Standardization (AEOLUS) [1]. AEOLUS aims to clean and normalize FAERS data by removing duplicate cases. This is done by mapping to standardized vocabularies and ontologies: RxNorm for drug names and ingredients, and SNOMED-CT for adverse event outcomes. The AEOLUS dataset is publicly available.

The algorithm used to calculate the results present in the nSides gateway is similar to the previously developed OFFSIDES and TWOSIDES databases, which used raw FAERS data [2] instead of AEOLUS. A standard signal detection algorithm involves conducting a disproportionality analysis by comparing the observed reporting frequency of a drug / drug combination and outcome to the expected reporting frequency of all other drugs / drug combinations and the same outcome. This metric is known as the Proportional Reporting Ratio (PRR) of the outcome for a particular drug or drug combination. If the outcome occurred by chance, the frequencies will be equal and the PRR will be one. If the PRR is much larger than one, we reject the null hypothesis. To reduce sampling variance and selection bias, we implemented propensity score matching on the FAERS data to form the groups used in the disproportionality analysis. This procedure—known as SCRUB—matches cases and controls between patients exposed and not exposed to a particular drug (OFFSIDES) or two drugs (TWOSIDES) to mitigate confounding biases. nSides extends this methodology to an arbitrary number of drugs, which can be specified by the user.

Presented at Gateways 2017, University of Michigan, Ann Arbor, MI, October 23-25, 2017.
https://gateways2017.figshare.com/

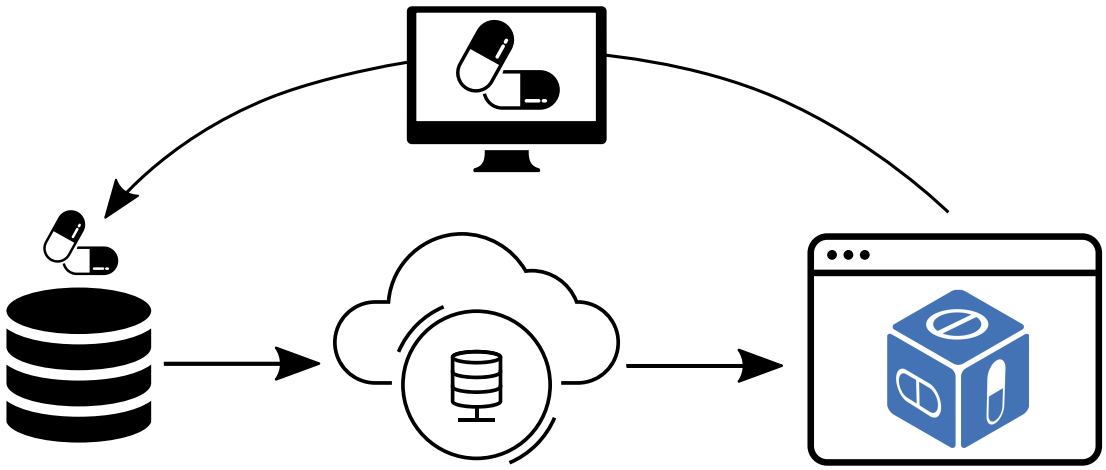

Fig. 1. A schematic representation of the infrastructural components that comprise nSides. FAERS/AEOLUS data is processed on grid computing systems to generate an adverse event database (lower left) with a web front-end (lower right), which has an on-demand interface (top center) to request models for new drug combination adverse reactions. User queries are handled by a set of middleware tools and an Agave API (bottom center).

III. NSIDES RESULTS

The central purpose of nSides is to present the results of the disproportionality analyses over a wide range of drugs and side effects. There are over 4,000 drug products and 7,000 outcomes present in the nSides database. The gateway makes the database interactive by allowing the user to search for a single drug or combination of drugs. The top 10 effects associated with the drug(s) (ordered by number of reports in AEOLUS) are presented to the user. The user is also able to retrieve results for other effects. The drug names are mapped to RxNorm identifiers and a remote MongoDB database instance is searched. Effects are returned in the form of 2 plots, showing the PRR and the number of cumulative reports (respectively) in the AEOLUS dataset as functions of the report year. We have populated the nSides database with drug effect results for all single drugs, as well as a number of drug pairs. If results for a combination of n drugs are not present in the database, the user is able to submit a request to have the results computed remotely, as we describe below in §V. A schematic of the different components of nSides are shown in Figure 1 and described in the sections below.

IV. DRUG INTERACTIONS

Users can search nSides for outcomes that are statistically associated to 2 or more interacting drugs. When users search for multiple drugs, results of the disproportionality analysis are presented for the drugs in combination. An example is shown in Figure 2. Additionally, lists of drugs in the same RxNorm classes of each of the drugs searched will appear in order of the number of reports present in FAERS. Effectively, it is possible to see the change in PRR caused by ”swapping” one of the drugs by another in the same RxNorm class. Drug safety – including the possibility of reducing adverse outcomes from combinations of drugs – can be assessed in a data-driven way.

V. CUSTOM DRUG MODELS

There are over 4,000 drugs available for analysis in nSides. It is not feasible to compute models for every combination of 2 more drugs. There are ≈250,000 drug combinations involving 2 drugs with enough reports feasible to generate drug effects and ≈8,000,000 involving combinations 3 drugs. The nSides database is only initially populated with combinations involving a small collection of drugs. If a combination of drugs is searched that does not have effect results present in the database, the user is given the option to request the computation of a model.

Generating a model is computationally intensive and can take many hours. To optimize this process, we use the Open Science Grid to generate models. When a user requests a model, a job request is submitted to run on the Open Science Grid. When completed, the result is added to the nSides database and is made available for all users. This procedure allows nSides to become populated in a user-driven way.

VI. FRONT-END FRAMEWORK

The nSides web application front-end is implemented using JavaScript (React and D3.js) (Figure 2). We chose these frameworks for their speed and reliability across all modern web browsers. Additionally, we designed the web application to be responsive to screen sizes of both desktop and mobile devices. We also provide an option for users to download the entire database or subsets of interest as CSV files. For job submission to the Open Science Grid, we designed and implemented a user interface to run jobs for custom drug models selected by users.

VII. BACK-END FRAMEWORK

The back-end to the nSides app is implemented in Python (using the Flask web framework), including a fully-featured REST API for querying the remote MongoDB database instance containing computed model results and drug/outcome

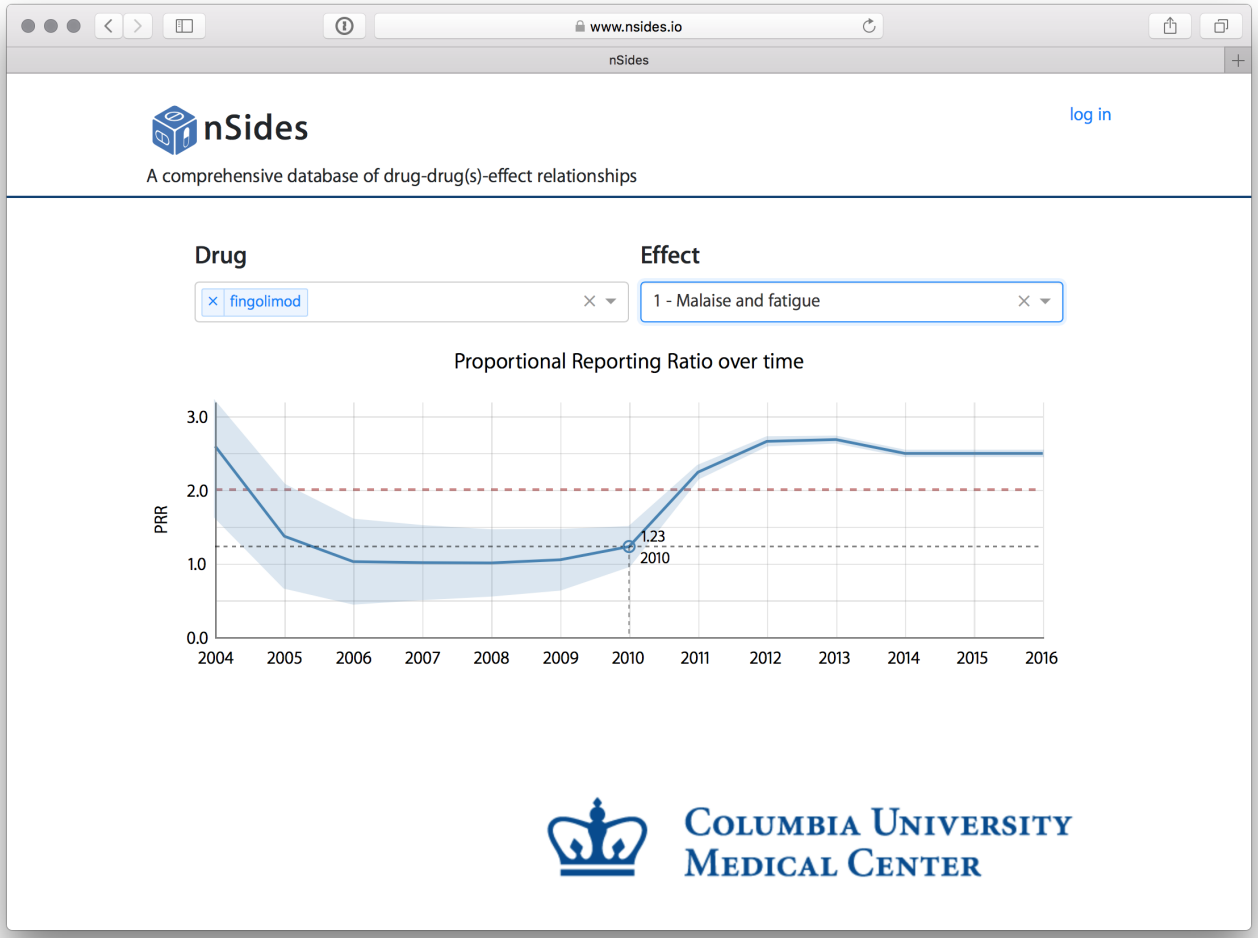

Fig. 2. Screenshot of nSides gateway showing PRR vs. year for "Malaise and fatigue" in response to the immunosuppressive drug "fingolimod", a well-known side effect and the first result returned by nSides.

vocabulary mappings. The nSides web application and back-end databases are each hosted on Amazon Web Services EC2 virtual Linux servers.

The sequential job dependencies of SCRUB require the use of the Directed Acyclic Graph Manager (DAGMan) within condor, which automatically spawns jobs once all of their dependencies have completed execution. Because of this constraint, we created and implemented a dedicated job submission API template in Agave [5], which enables users to submit jobs for new drug combinations directly to the Open Science Grid. Various data and job lifecycle components (including archiving results) are also handled by the Agave API. When user-submitted jobs complete on the Open Science Grid resources, the Agave API populates the results into the aforementioned MongoDB instance, so that the results are automatically available for visualization. The Agave API management platform is fully integrated into the nSides web front-end, which includes secure user authentication and a dashboard for visualizing the status of submitted jobs.

VIII. DISCUSSION

nSides is the latest output of a series of data-driven studies that aim to make use of public data and statistical modeling to discover adverse and off-label side effects. Given the scope of nSides and its ability to assist in many areas of pharmacovigilance for multiple stakeholders (researchers, clinicians, educators, and even patients), we aim to maintain nSides as a standalone, open-source, and gratis public resource for the foreseeable future.

IX. CONCLUSIONS

The nSides gateway allows researchers and the general public to mine FAERS for statistically significant drug effects in an interactive way. nSides has the potential to improve drug safety by providing alternate drug recommendations to reduce adverse outcomes that are associated to a drug interaction.

The nSides source (including the front-end web app, back-end distributed computing framework, and middleware) is available on GitHub:
https://github.com/tatonetti-lab/nsides

ACKNOWLEDGMENT

R.S.V., J.D.R., T.L., V.N., and N.P.T. are supported by grants from the National Institutes of Health: NIGMS R01GM107145 and NCATS OT3TR002027. R.S.V. and N.P.T. are also supported by the Herbert Irving Fellowship. This research is conducted using resources provided by the Open Science Grid [3], [4], which is supported by the National Science Foundation award 1148698, and the U.S. Department of Energy's Office of Science. This research is also supported by the Science Gateways Community Institute's Extended Developer Support program, funded by the National Science Foundation award 1547611.

REFERENCES

[1] Banda, J. M. et al. A curated and standardized adverse drug event resource to accelerate drug safety research. Sci. Data 3:160026 doi: 10.1038/sdata.2016.26 (2016).

[2] Tatonetti, N. P., Ye, P. P., Daneshjou, R., & Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. Science Translational Medicine, 4(125), 125ra31. Retrieved from http://stm.sciencemag.org/content/4/125/125ra31.abstract

[3] Pordes, R. et al. (2007). "The Open Science Grid", J. Phys. Conf. Ser. 78, 012057.doi:10.1088/1742-6596/78/1/012057.

[4] Sfiligoi, I., Bradley, D. C., Holzman, B., Mhashilkar, P., Padhi, S. and Wurthwein, F. (2009). "The Pilot Way to Grid Resources Using glideinWMS", 2009 WRI World Congress on Computer Science and Information Engineering, Vol. 2, pp. 428432. doi:10.1109/CSIE.2009.950.

[5] Agave API Platform, https://agaveapi.co/.

[6] Kuhn, Michael, et al. "The SIDER database of drugs and side effects." Nucleic acids research 44.D1 (2015): D1075-D1079.