Big Data Mining and Adverse Event Pattern Analysis in Clinical Drug Trials

Callie Federer,[1,2,] Minjae Yoo,[1,*] and Aik Choon Tan[1,3,4]*

[1]Translational Bioinformatics and Cancer Systems Biology Laboratory, Division of Medical Oncology, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado.
[2]Computational Bioscience Graduate Program, Department of Pharmacology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado.
[3]Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, Colorado.
[4]University of Colorado Cancer Center, University of Colorado Anschutz Medical Campus, Aurora, Colorado.
**These authors contributed equally to this research.*

ABSTRACT

Drug adverse events (AEs) are a major health threat to patients seeking medical treatment and a significant barrier in drug discovery and development. AEs are now required to be submitted during clinical trials and can be extracted from ClinicalTrials.gov (https://clinicaltrials.gov/), a database of clinical studies around the world. By extracting drug and AE information from ClinicalTrials.gov and structuring it into a database, drug-AEs could be established for future drug development and repositioning. To our knowledge, current AE databases contain mainly U.S. Food and Drug Administration (FDA)-approved drugs. However, our database contains both FDA-approved and experimental compounds extracted from ClinicalTrials.gov. Our database contains 8,161 clinical trials of 3,102,675 patients and 713,103 reported AEs. We extracted the information from ClinicalTrials.gov using a set of python scripts, and then used regular expressions and a drug dictionary to process and structure relevant information into a relational database. We performed data mining and pattern analysis of drug-AEs in our database. Our database can serve as a tool to assist researchers to discover drug-AE relationships for developing, repositioning, and repurposing drugs.

Keywords: adverse events, big data mining, pattern analysis, clinical drug trials, bioinformatics

INTRODUCTION

Adverse events (AEs) are unintended and undesirable effects as a result of the use of drug treatment or other medical product in a patient. AEs represent a significant barrier in drug development for patient treatment. Serious drug adverse effect is the fourth leading cause of death in the United States, with over 100,000 people dying from this each year.[1] Approximately 30% of failures in drug clinical trials are due to the intensity of adverse side effects.[2] We hypothesize that learning about drug-related AEs from clinical trial data will provide new insights and reveal unexpected relationships between drugs, AEs, and drug targets for future drug development and biomedical research.

Recently, many research efforts have been focused on understanding the relationships between drug targets and AEs, with the aim to elucidate the molecular mechanisms underlying drug adverse effects for better development and repurposing of drugs. Two important data sources—drug target and AE databases (AEDB)—are needed to investigate the relationships between drugs, drug targets, and AEs. There are many databases and repositories of drugs and target relationships, such as PubChem,[3] ChemBank,[4] DrugBank,[5] BindingDB,[6] and DSigDB.[7] The U.S. Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) represents one of the most popular AEDB of FDA-approved drugs. FAERS is designed to support the FDA's postmarketing safety surveillance program for drugs and therapeutic products. FAERS collects AEs and medication error reports that manufacturers, healthcare professionals, and consumers submit to the FDA. Another popular source for extracting drug AEs is FDA package inserts for products.

Computational tools and approaches such as machine learning and text mining have been developed and used to construct resources and predictors of drug-AE relationships. SIDER,[8] the side effect resource, represents one of the earliest computational approaches to extracting drug AEs from FDA package inserts. Currently, SIDER contains 5,868 AEs, 1,430 drugs, and 139,756 drug-AE pairs.[9] In contrast, the

OFFSIDES database contains AEs not listed on the FDA's official drug label. Currently, this database contains 1,332 drugs, 10,097 AEs, and 438,801 drug-AE pairs, and these information could be used for polypharmacology research.[10] MetaADEDB is a database of adverse drug events and has been recently developed by combining SIDER, OFFSIDES, and the Comparative Toxicology Database.[11] Machine learning approaches have been used to build predictors for drug-AEs, by using properties such as chemical structures, drug targets, structural information, and drug–protein interaction networks.[12–15]

Although research tools and resources have been developed for predicting drug-AE relationships, knowledge and prediction of drug-AEs are still far from perfect. One of the main limitations of the current approaches is that they are limited to FDA-approved drugs,[9,10] as the primary source of AEs. Other experimental compounds tested in clinical trials were not captured by the current approaches. These experimental compound AEs in clinical trials represent an untapped resource, and could potentially provide new knowledge of drug-AE relationships. Collections of trials data and results are now available in various clinical trials registries, such as ClinicalTrials.gov. ClinicalTrial.gov provides completed results for approved and experimental drugs tested in multiple medical conditions across various trial phases. Thus, ClinicalTrials.gov offers a unique opportunity to perform unbiased exploration and learning of drug-AE relationships.

To understand the patterns of AEs reported in clinical trials, we performed "big data mining" on the published results from the ClinicalTrials.gov website. We developed and implemented data extraction and text mining programs to automatically retrieve clinical trials with AEs. We developed a novel AE relational database that links between clinical trials, drugs, and AEs. We parsed the data extracted from the ClinicalTrials.gov website into our database. Then, we performed data mining, pattern analysis, and data visualization on the reported AEs. To illustrate one application of our database, we utilized the proportional reporting ratio (PRR) for comparing selected small molecule kinase inhibitors and AEs.

MATERIALS AND METHODS

ClinicalTrials.gov and HTML Contents

ClinicalTrials.gov is a registry and results database of both publicly and privately supported clinical studies of human participants around the world (https://clinicaltrials.gov). The goal of a clinical trial is to determine if an intervention, in comparison to other available interventions (or no intervention), will be helpful for a condition and safe for patients. The ClinicalTrials.gov results database was launched in September 2008 after the FDA Amendments Act of 2007 required the submission of basic results of certain clinical trials. In 2009, the submission of AE information was required. In this study, we focus on clinical trials where the intervention is a drug or drug combinations.

Data Extraction and Text Mining

To extract AEs and other related trial information from the ClinicalTrials.gov website, we implemented a Python script that utilizes Beautiful Soup.[16] Beautiful Soup is a Python library for navigating and searching a parse tree, which is used in concordance with the html5lib Python parser for extracting data from websites. The list of Clinical Trial IDs included in this database version is from October 5, 2015 or earlier. We also developed a set of Python scripts for postprocessing and uploading the data into MySQL tables. The postprocessing steps include regular expressions, text mining, and a dictionary-based approach for extracting certain fields as information was entered in various formats in the ClinicalTrials.gov published results. We used MySQL version 5.7.11, which is an open source database. To insert the data into the MySQL tables, we used PyMySQL, a pure Python MySQL driver compatible with Python 3, which allows the execution of MySQL statements. Using the Python scripts, we extracted the cohort information from the Reporting Group table. We grouped both the Serious AEs table and Other AEs table, which included the AE name, category, the number of patients affected in the cohort, and the total number of patients in the cohort. The workflow of this data extraction and text mining is illustrated in *Figure 1A*.

Drug List

We extracted drug information from the clinical trial results using a drug dictionary. This not only results in fewer false-positive drug identifications in the cohort descriptions but also restricts drug and AE relationships to a list of known drug names for further analysis. In this study, we used the FDA-approved drugs and experimental compounds obtained from DSigDB.[7] DSigDB currently holds 17,389 unique compounds and 19,531 drug target genes, and is freely available at http://tanlab.ucdenver.edu/DSigDB

AE Database, Analysis and Visualization

The extracted drugs and AEs were uploaded to our adverse event database (AEDB). Our AEDB contains nine tables, focusing on drugs, AEs, cohorts, and drug targets. The AEDB schema is illustrated in *Figure 2*. The AEDB was developed using the open source database MySQL.

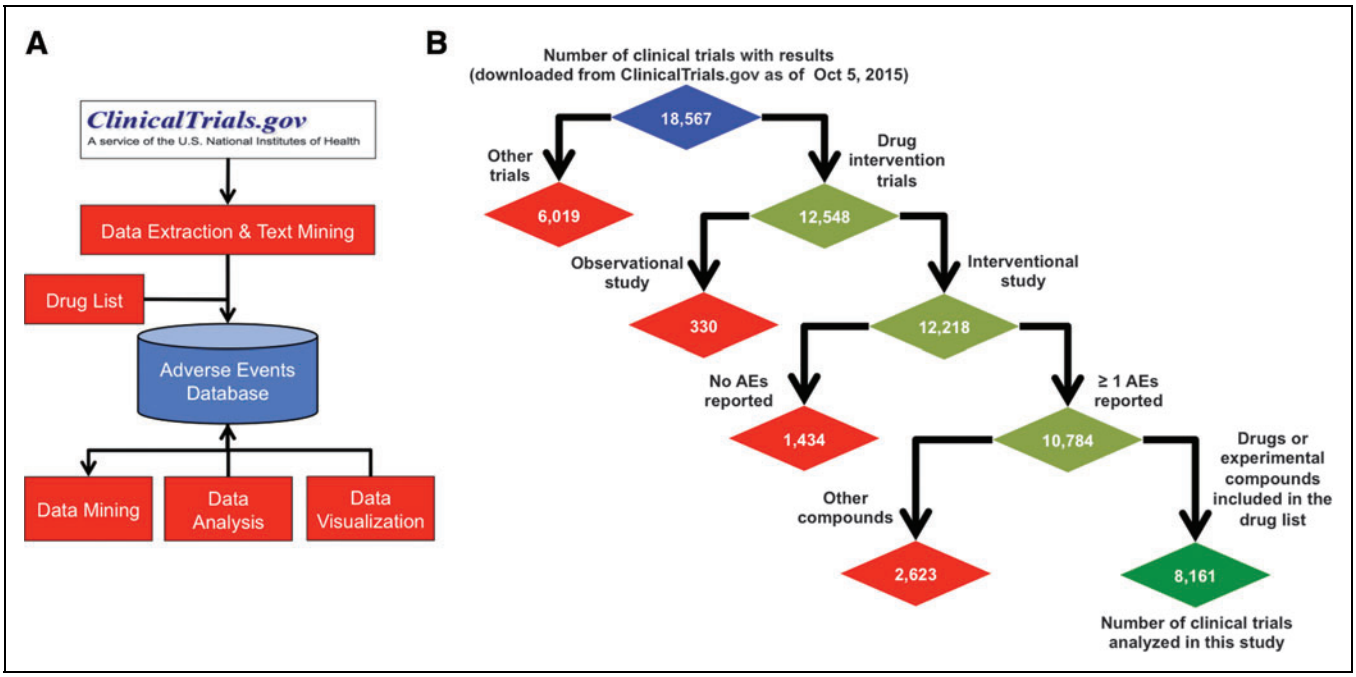

Fig. 1. Overview of the research strategy. **(A)** The workflow for extracting and data mining clinical drug trial data from ClinicalTrials.gov. **(B)** The flow chart of filtering clinical trials used in this study. Color images available online at www.liebertpub.com/adt

PRR Analysis

We used the PRR that summarizes the extent to which a certain AE is reported for patients taking a particular drug compared with the frequency at which the same AE is reported in other drugs.[17] The PRR has been used to find signal in AEs for safety reporting in drugs.[10,17,18] A PRR greater than one implies that the drug of interest had a higher reported frequency of the AE than the rest of the drugs.

RESULTS

Summary of the AEs Extracted from ClinicalTrials.gov1

We downloaded 18,567 trials with results reported in the ClinicalTrials.gov database as of October 5, 2015. The workflow of extracting the drug AEs for this study is illustrated in *Figure 1B*. We selected clinical trials that had "Drug" in the Intervention type. This generates a list of 12,548 "Drug" trials

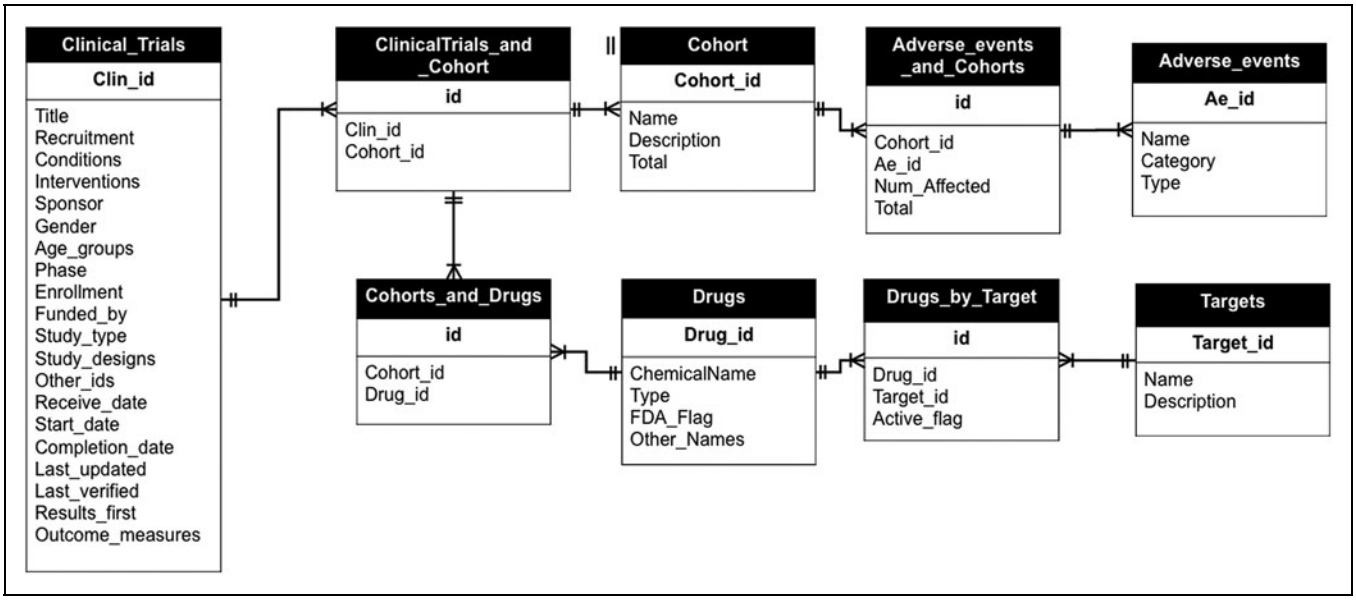

Fig. 2. Entity-relationship model of the AEDB. AEDB, adverse event database.

from the total number of trials in the results list. Then, we filtered 330 observational studies, and we further filtered 1,434 trials with no AEs reported in the results. We matched the drugs from the 10,784 trials that have at least one reported AE using a list of FDA-approved drugs, small molecules, and monoclonal antibodies with known targets. Using this drug list, we extracted drug-AE relationships from 8,161 trials. We also extracted placebo-related AEs from this list of trials. *Table 1* provides some examples of data extracted from these trials.

Table 2 summarizes the statistics of the data in our AEDB. We have extracted 8,161 trials from ClinicalTrials.gov, in which more than 3 million patients participated. Among the 1,248 drugs that were extracted from these trials, 634 were FDA-approved drugs. Placebo was used in 3,404 trials, representing 42% of the clinical trials in this study. The 3 million patients were tested in 20,739 cohorts across these trials. A total number of 31,267 AEs were extracted from these trials that span across 26 AE categories. A total number of 713,103 AEs are reported in our study. There are more than 3,000 medical conditions tested in these trials. The AEDB contains a unique data set collected from clinical trials, which provide an opportunity to study drug-AE relationships. The AEDB data set is different from the other existing databases based on FAERS. *Supplementary Figure S1* (Supplementary Data are available online at www.liebertpub.com/adt) shows a Venn diagram of drugs in AEDB, SIDER, and OFFSIDES. As shown, AEDB has 539 drugs or experimental compounds that are currently not included in SIDER or OFFSIDES. *Supplementary Figure S2* illustrates the summary statistics of drugs and AEs in AEDB.

Table 1. Data Extracted from ClinicalTrials.gov

Data Extracted	Data Examples
Clinical trials	NCT00860743
	NCT00718770
	NCT01909141
Drugs	Bevacizumab
	Carboplatin
	Metformin
Cohorts	"Cohort 1, healthy adults"
	"Cohort 1, adolescents 12 to 17 years old, 4060 mg"
	"Cohort 2, children 6 to 12 years old, 2060 mg"
Adverse event names	"Fever"
	"Vomiting"
	"Nausea"
Adverse event categories	"Nervous system disorders"
	"Vascular disorders"
Conditions	Diabetes mellitus
	Alzheimer's disease
	Prostate cancer

Table 2. Summary Statistics of the Database

Description	Counts
Number of clinical trials	8,161
Number of patients	3,102,675
Number of drugs	1,248
Number of FDA-approved drugs	634
Number of non-FDA-approved drugs	614
Number of cohorts	20,739
Number of adverse event names	31,267
Number of adverse event categories	26
Number of reported adverse events	713,103
Number of conditions	3,279

FDA, U.S. Food and Drug Administration.

Statistics of the AEs

We extracted and grouped the AEs from ClinicalTrials.gov into 26 unique AE categories (*Fig. 3A*). The top three most common AE categories were gastrointestinal disorders, infections and infestations, and nervous system disorders. The least common AE category was social circumstances (*Fig. 3A*). Although the AE names reported across the clinical trials did not follow a dictionary, the 26 unique AE categories comply with the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE v4.0). The 10 most common AEs found in these clinical trials were headache, nausea, dizziness, vomiting, fatigue, constipation, diarrhea, back pain, nasopharyngitis (common cold), and cough (*Fig. 3B*). These AEs were found in various medical conditions. We observed an increased number of AEs per patient reported in recent years, potentially due to the requirement to report AE results in the ClinicalTrials.gov repository (*Supplementary Fig. S3*).

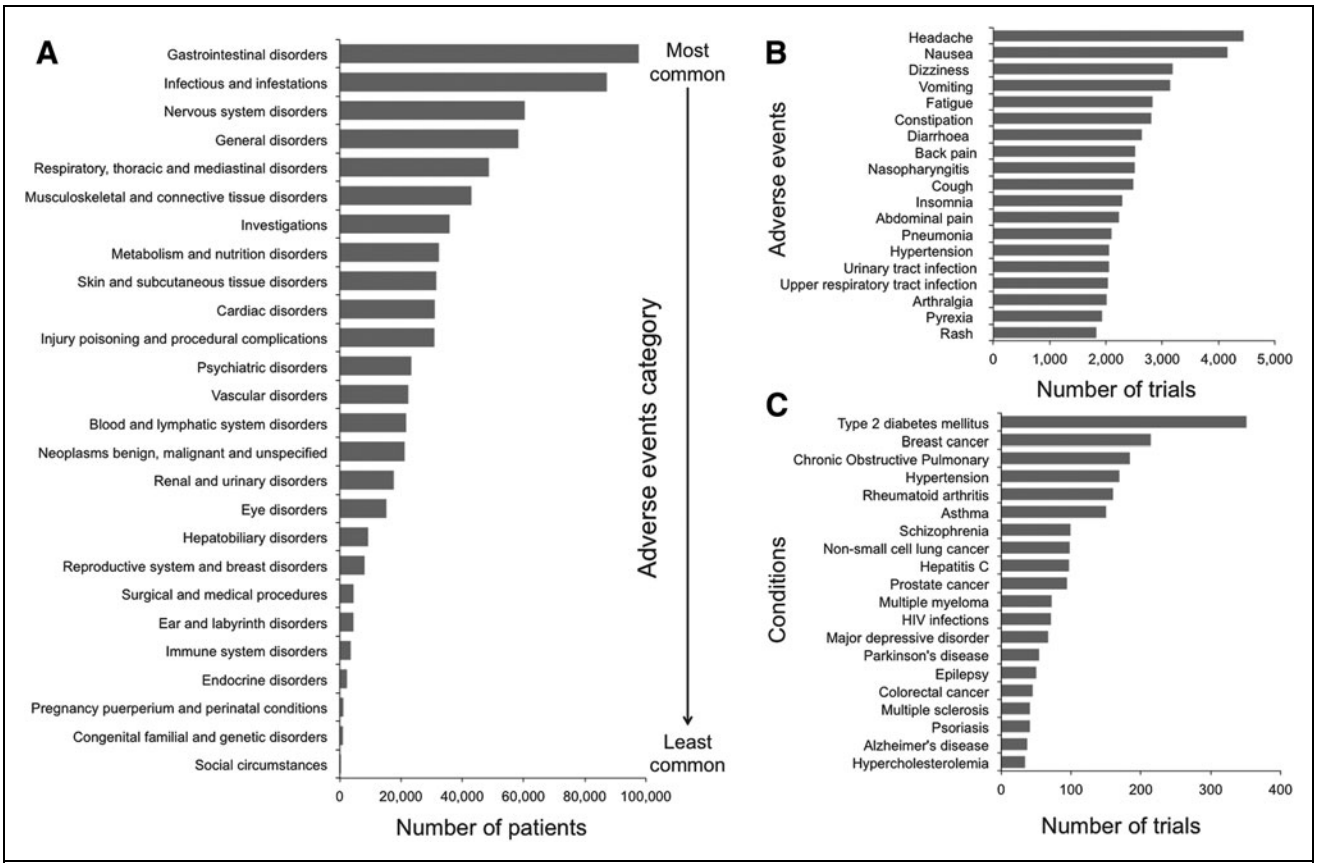

Fig. 3. Statistics of the AEs. **(A)** The 26 AE categories collected in AEDB. **(B)** The top 20 most common AEs reported in AEDB. **(C)** The top 20 most common medical conditions in AEDB.

The 10 most common conditions investigated in these clinical trials were Type 2 diabetes mellitus, breast cancer, chronic obstructive pulmonary disease, hypertension, rheumatoid arthritis, asthma, schizophrenia, nonsmall cell lung cancer, hepatitis C, and prostate cancer (*Fig. 3C*). To explore the disease-AE relationships, we performed PRR analysis on the top 20 conditions. *Supplementary Figure S4* shows the disease-AE relationships in a heatmap. For example, auditory hallucination is strongly correlated with schizophrenia, major depressive disorder, Parkinson's disease, epilepsy, and Alzheimer's disease.

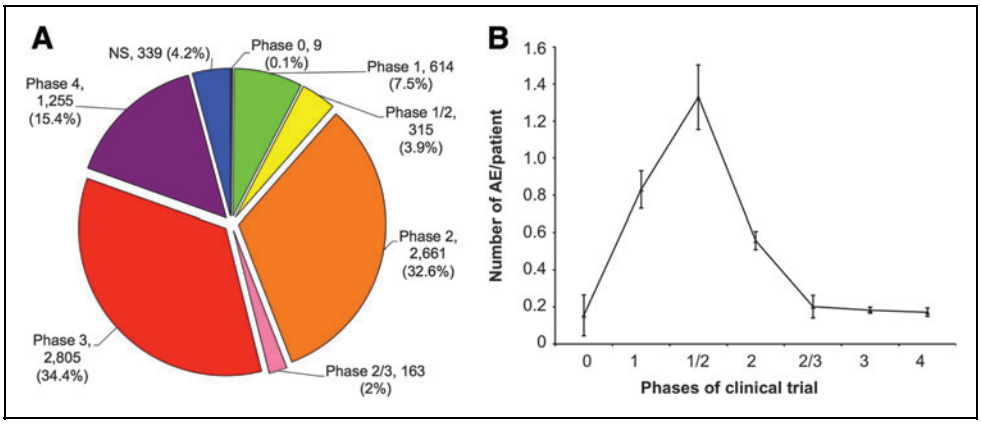

Fig. 4. AEs in different phases of clinical trials. **(A)** Distribution of the different phases of clinical trials. **(B)** Average number of AEs per patient in different phases of clinical trials. N.S., not specified. Error bar represents the standard error of the mean. Color images available online at www.liebertpub.com/adt

AEs in Different Phases of Clinical Trials

Next, we investigate the AEs recorded in the different phases of clinical trials. *Figure 4A* shows the breakdown of the clinical trial phases in this study. The top three phases with the most complete AE results were Phase 3 (34.4%), Phase 2 (32.6%), and Phase 4 (15.4%). We found that Phase 1/2 patients experienced the highest number of AEs, followed by Phase 1 and Phase 2 patients. This is not surprising as these early trial

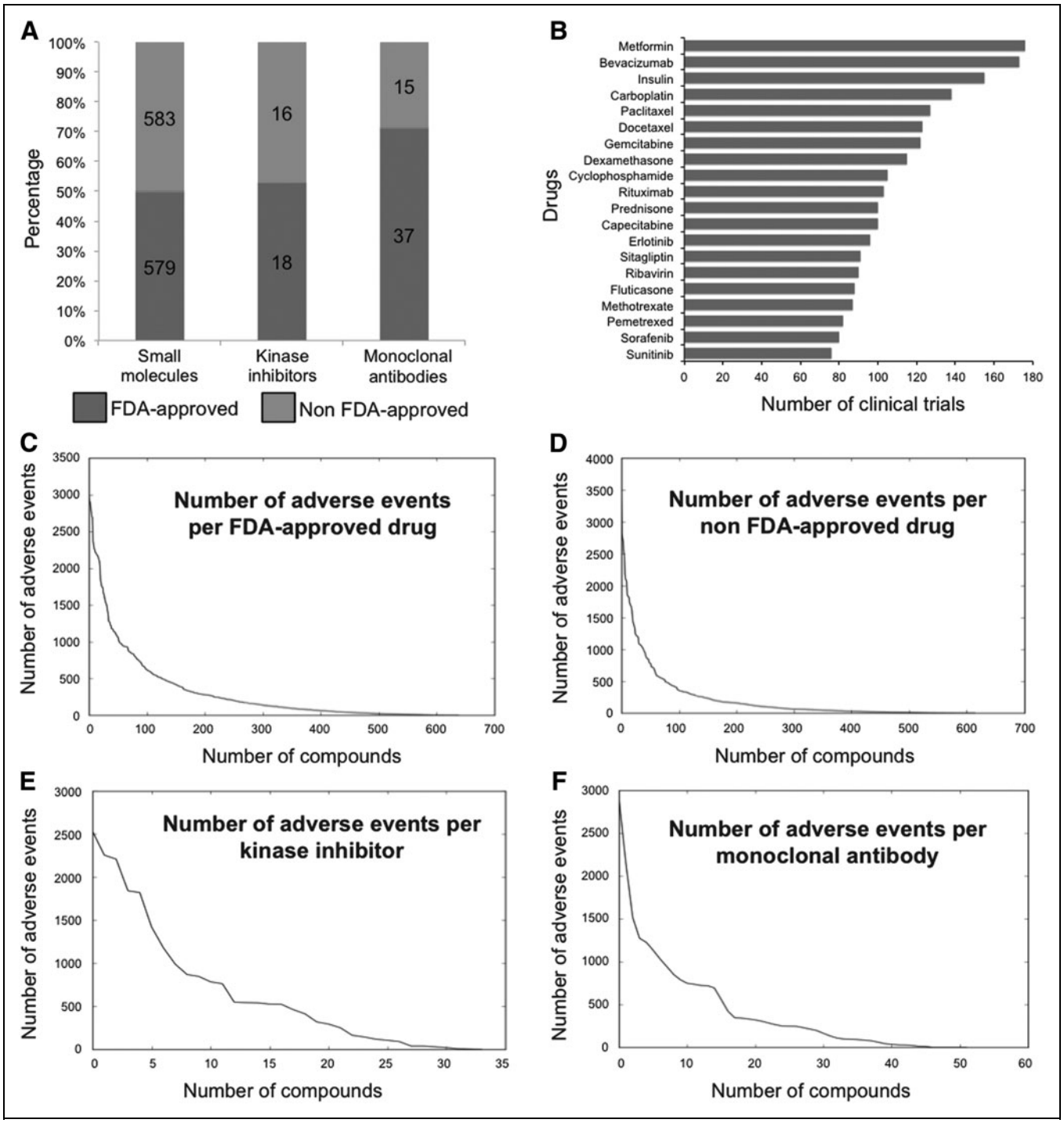

Fig. 5. Drug-AE relationships. **(A)** Distribution of the drug classes collected in AEDB. **(B)** The top 20 most common drugs reported in AEDB. **(C)** The number of AEs per FDA-approved drug. **(D)** The number of AEs per non-FDA-approved drug. **(E)** The number of AEs per kinase inhibitor. **(F)** The number of AEs per monoclonal antibody. FDA, U.S. Food and Drug Administration.

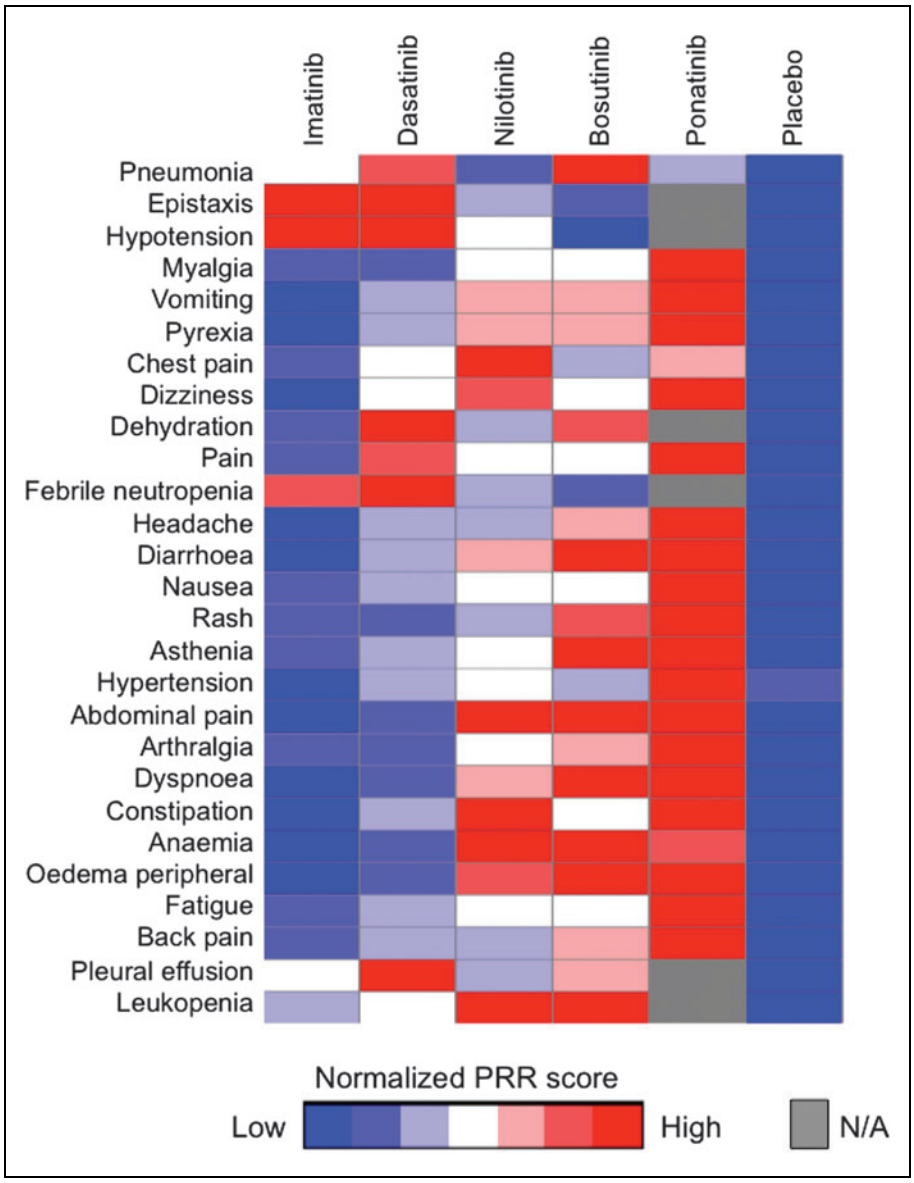

Fig. 6. Kinase inhibitor-AE relationships. Heatmap of the PRR of the top 10 AEs of imatinib, dasatinib, nilotinib, bosutinib, ponatinib, and placebo. The PRR is normalized per AE, where *red* and *blue* colors indicate high and low frequencies, respectively. PRR, proportional reporting ratio. Color images available online at www.liebertpub.com/adt

phases are enriched with experimental compounds, and the main objective of these trials is to determine the toxicity of these compounds in patients. Accordingly, Phase 3 and 4 patients experienced the least number of AEs; these trials are late-stage trials where the main objectives are the efficacy of the drugs (Phase 3) and post-marketing surveillance of the drugs (Phase 4). *Figure 4B* shows the average number of AEs per patient in these different phases of trials.

Statistics of the Drug-AE Relationships

Next, we investigate the drug-AE relationships that were extracted from ClinicalTrials.gov and included in our database. Using the drug list that we compiled, we extracted 1,248 drugs from the 8,161 clinical trials. Out of these 1,248 drugs, 634 were FDA-approved drugs. Among the FDA-approved drugs, 18 were kinase inhibitors, 37 were monoclonal antibodies, and the remaining were small molecule drugs. In the non-FDA-approved drugs category, 16 were kinase inhibitors, 15 were monoclonal antibodies, and 583 were other small molecules. Among the 8,161 clinical trials collected in our database, 5,981 contained experimental compounds from the trials. This demonstrates that there are a significant number of experimental compounds published with AEs that are not included in the databases that contain only FDA-approved drugs. *Figure 5A* summarizes the distributions of the drugs in AEDB. Metformin, an FDA-approved drug for the treatment of type 2 diabetes mellitus, is the most commonly used drug in these clinical trials. The other top 10 commonly used drugs were bevacizumab, insulin, carboplatin, paclitaxel, docetaxel, gemcitabine, dexamethasone, cyclophosphamide, and rituximab. *Figure 5B* shows the top 20 commonly used drugs in AEDB. *Figure 5C* and *D* illustrate the number of AEs due to FDA-approved and non-FDA-approved drugs, respectively. *Figure 5E* and *F* show the number of AEs associated with kinase inhibitors and monoclonal antibodies, respectively. We found that kinase inhibitors have a slightly higher number of AEs, compared with monoclonal antibodies, suggesting that kinase inhibitors might have more "off-targets."

Case Study: Kinase Inhibitor-AE Relationships

To illustrate one application of AEDB, we performed PRR for comparing selected small molecule kinase inhibitors and AEs. Protein kinases play a key role as regulators and transducers of signaling in eukaryotic cells, and represent the largest and well-studied "druggable" families in the human

genome.[19] Many kinases are mutated in cancer genomes, and cancer cells depend on these mutated kinases for proliferation, growth, and survival signaling. Therefore, small molecule inhibitors that inhibit kinases either in wild-type or mutated forms are actively studied in the pharmaceutical industry and academia. However, due to the conserved sequence similarity between kinases, many kinase inhibitors have off-target effects, which can ultimately lead to AEs in patients.

The majority of chronic myelogenous leukemia (CML) cases are driven by the oncogenic kinase fusion of *BCR-ABL*. Imatinib is a small molecule kinase inhibitor that specifically inhibits the activity of *BCR-ABL*, and dramatically improves the survival of CML patients.[20] Imatinib represents the first FDA-approved kinase inhibitor in treating CML; additional four kinase inhibitors (dasatinib, nilotinib, bosutinib, and ponatinib) are also approved by FDA for this disease. However, these newer kinase inhibitors cause more serious AEs. To study these AEs, we focused on the five kinase inhibitors approved for the treatment of CML: imatinib, nilotinib, dasatinib, bosutinib, and ponatinib. We used PRR to evaluate AEs reported for each of the kinase inhibitors. *Figure 6* shows the top 10 AEs found in each kinase inhibitor in the database compared to placebo. From this heatmap, it is clear that ponatinib has more AEs and a different AE profile compared to the other kinase inhibitors.

We further investigated the selected vascular-related AEs associated with these kinase inhibitors. These vascular AEs have emerged as a serious consequence of the treatment of kinase inhibitors.[21,22] From the analysis, we found that these kinase inhibitors have a higher PRR score in peripheral arterial occlusive disease, embolism, hypertension, platelet dysfunction, hyperglycemia, and hair loss, compared with the other drugs in the database (Table 3). Specifically, ponatinib has the highest PRR score in peripheral arterial occlusive disease. This finding is supported by the FDA warnings for ponatinib, which include serious AEs related to life-threatening blood clots and severe narrowing of the blood vessels. As a result, the FDA issued a temporary marketing suspension of ponatinib in October 2013 and began to require extra safety measures for ponatinib in December 2013, before the company resumed marketing.[23,24] This suggests that performing data mining on our database may reveal new knowledge about drug-AE relationships.

Table 3. Vascular Event Proportional Reporting Ratios for the Five Kinase Inhibitors Commonly Used to Treat Chronic Myelogenous Leukemia Patients

	Vascular Adverse Events						
Kinase Inhibitors	Peripheral Arterial Occlusive Disease	Embolism	Hypertension	Platelet Dysfunction	Hyperglycemia	Hair Loss Alopecia	Vascular Disorders
Imatinib	7.416	4.874	4.550	10.929	4.944	4.398	5.481
Dasatinib	NA	2.959	8.161	17.624	10.720	4.427	4.263
Nilotinib	31.497	2.070	10.541	11.604	14.998	4.239	4.810
Bosutinib	NA	5.457	7.719	10.197	5.272	4.443	3.719
Ponatinib	374.810	NA	41.811	69.044	NA	7.486	9.158
Placebo	2.065	1.861	1.957	0.326	1.404	0.000	1.836

NA, not applicable due to no data.

DISCUSSION

We have performed "big data" mining and pattern analysis of drug AEs in ClinicalTrials.gov. We extracted drug-AE relationships from 8,161 clinical trials, in which more than 3 million individuals participated. A total of 1,248 drugs and a total of 31,267 AEs were extracted from these trials. The AEs extracted from these trials span across 26 AE categories. To facilitate data analysis, we have developed AEDB to store and manage the drugs and AEs extracted from ClinicalTrials.gov. We performed data mining and analyzing drug-AE relationships using AEDB.

Current drug-AEDB such as SIDER focus on only FDA-approved drugs, as most of the AEs are obtained from the FAERS or FDA drug labels. In contrast, AEDB extracted both FDA-approved drugs and experimental compounds from clinical trial data. These experimental compounds and AE relationships have not been fully studied, and are an untapped resource for mining new drug-AE relationships. We believe that our database provides a unique opportunity to learn and extract drug-AE relationships, and it is complementary to the existing AE resources. Our database can easily be scaled up to capture new data deposited to ClinicalTrials.gov. We plan to periodically update the database with new results from ClinicalTrials.gov.

Trial data reported in ClinicalTrials.gov are a new source of big data for biomedical research, as ClinicalTrials.gov currently holds more than 217,000 studies in its registry. However, several challenges exist in extracting data from this data source. First, the lack of standards in clinical trial data elements makes it more difficult to extract and map data. Second, no standard drug names or dictionary were used in the repository, and manual inspection is required to correctly map the drug names. Third, different ontologies were used in reporting AEs, and manual inspection is required to map and consolidate the name of AEs. Finally, typos in data entry require manual inspection and correction. Two recent studies have explored the data in ClinicalTrials.gov. One study focuses on learning disease relationships from clinical drug trials,[25] and the other focuses on extracting genetic alteration information for personalized cancer therapy.[26] Both studies used text mining approaches to extract the trial information from ClinicalTrials.gov.

In the future, we would like to investigate the drug target-AE relationships in our database to elucidate the molecular mechanisms of drug actions and improve personalized medicine using previously published methodologies.[12–15,27] We would like to use AEs for predicting novel drug–target interactions for drug repurposing and repositioning. We would also like to develop an interactive web portal such as[8,9,28,29] that users can utilize to query, retrieve, and analyze data collected in this database. The database would be searchable by drug, drug target, AE, condition, and/or clinical trial. Our study has the limitation of not considering the different drug dosages and their related AEs, which we plan to address in our future work.

In conclusion, we have extracted clinical drug trial data from ClinicalTrials.gov and structured it into a database for mining, predicting, and visualizing AEs. We developed and implemented data extraction and text mining programs to automatically retrieve clinical trials with AEs. We developed a novel AE relational database that links between clinical trials, drugs, and AEs. We illustrated the application of the database by performing the PRRs for comparing the AEs of five common kinase inhibitors with those of other drugs in the database. We found that the signal in the AEs with higher frequencies and the results were corroborated by published studies. Our database can serve as a tool to assist researchers discover drug-AE relationships for developing, repositioning, and repurposing drugs.

ACKNOWLEDGMENTS

We would like to acknowledge the Tan Lab members for their constructive comments on this project. We thank Susan Kim for suggestions and editing of the article. This work is partly supported by the National Institutes of Health P50CA058187, P30CA046934, Cancer League of Colorado, and the David F. and Margaret T. Grohne Family Foundation.

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

1. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y: When good drugs go bad. *Nature* 2007;446:975–977.
2. Hopkins AL: Network pharmacology: The next paradigm in drug discovery. *Nature Chem Biol* 2008;4:682–690.
3. Kim S, Thiessen PA, Bolton EE, *et al.*: PubChem Substance and Compound databases. *Nucleic Acids Res* 2016;44:D1202–D1213.
4. Davies M, Nowotka M, Papadatos G, *et al.*: ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Res* 2015;43:W612–W620.
5. Law V, Knox C, Djoumbou Y, *et al.*: DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014;42:D1091–D1097.
6. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J: BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res* 2016;44:D1045–D1053.
7. Yoo M, Shin J, Kim J, *et al.*: DSigDB: Drug signatures database for gene set analysis. *Bioinformatics* 2015;31:3069–3071.
8. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P: A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010;6:343.
9. Kuhn M, Letunic I, Jensen LJ, Bork P: The SIDER database of drugs and side effects. *Nucleic Acids Res* 2016;44:D1075–D1079.
10. Tatonetti NP, Ye PP, Daneshjou R, Altman RB: Data-driven prediction of dug effects and interactions. *Sci Transl Med* 2012;4:125ra31.
11. Cheng F, Li W, Wang X, *et al.*: Adverse drug events: Database construction and in silico prediction. *J Chem Inf Model* 2013;53:744–752.
12. Bender A, Scheiber J, Glick M, *et al.*: Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* 2007;2:861–873.
13. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P: Drug target identification using side-effect similarity. *Science* 2008;321:263–266.
14. Liu M, Wu Y, Chen Y, *et al.*: Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc* 2012;19:e28–e35.
15. Mizutani S, Pauwels E, Stoven V, Goto S, Yamanishi Y: Relating drug-protein interaction network with drug side effects. *Bioinformatics* 2012;28:i522–i528.
16. Richardson L: BeautifulSoup. www.crummy.com/software/BeautifulSoup Web (Last accessed on August 31, 2016).
17. Evans S, Waller PC, Davis S: Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001;10:483–486.
18. Kass-Hout TA, Xu Z, Mohebbi M, *et al.*: OpenFDA: an innovative platform providing access to a wealth of FDA's publicly available data. *J Am Med Inform Assoc* 2016;23:596–600.
19. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: The protein kinase complement of the human genome. *Science* 2002;298:1912–1934.
20. Druker BJ, Talpaz M, Resta DJ, *et al.*: Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase inchronic myeloid leukemia. *N Engl J Med* 2001;344:1031–1037.

21. Pasvolsky O, Leader A, Iakobishvili Z, Wasserstrum Y, Kornowski R, Raanani P: Tyrosine kinase inhibitor associated vascular toxicity in chronic myeloid leukemia. *Cardiooncology* 2015;1:5.
22. Valent P, Hadzijusufovic E, Schernthaner GH, Wolf D, Rea D, le Coutre P: Vascular safety issues in CML patients treated with BCR/ABL1 kinase inhibitors. *Blood* 2015;125:901–906.
23. Tournaire G, Despas F, Huguet F, Montastruc JL, Bondon-Guitton E: Peripheral arterial occlusive disease during ponatinib therapy after failure of imatinib: A case report. *J Clin Pharm Ther* 2016;41:360–361.
24. FDA Drug Safety Communication: FDA asks manufacturer of the leukemia drug Iclusig (ponatinib) to suspend marketing and sales. www.fda.gov/Drugs/DrugSafety/ucm373040.htm (Last accessed on August 31, 2016)
25. Haslam B, Perez-Breva L: Learning disease relationships from clinical drug trials. *J Am Med Inform Assoc* 2016; DOI: http://dx.doi.org/10.1093/jamia/ocw003.
26. Xu J, Lee H-J, Zeng J, *et al.*: Extracting genetic alteration information for personalized cancer therapy from ClinicalTrials.gov. *J Am Med Inform Assoc* 2016;23:750–757.
27. Yang X, Huang Y, Crowson M, Li J, Maitland M, Lussier YA: Kinase inhibition-related adverse events predicted from in vitro kinome and clinical trial data. *J Biomed Inform* 2010;43:376–384.
28. Juan-Blanco T, Duran-Frigola M, Aloy P: IntSide: A web server for the chemical and biological examination of drug side effects. *Bioinformatics* 2015;31:612–613.
29. Liu X, Gao Y, Peng J, *et al.*: TarPred: A web application for predicting therapeutic and side effect targets of chemical compounds. *Bioinformatics* 2015;31:2049–2051.

Address correspondence to:
Aik Choon Tan, PhD
Translational Bioinformatics
and Cancer Systems Biology Laboratory
Division of Medical Oncology
Department of Medicine, School of Medicine
University of Colorado Anschutz Medical Campus
Mail Stop 8117
12801 E.17th Avenue, L18-8116
Aurora, CO 80045

E-mail: aikchoon.tan@ucdenver.edu

Abbreviations Used

AE = adverse event
AEDB = adverse event database
CML = chronic myelogenous leukemia
FAERS = FDA Adverse Event Reporting System
FDA = U.S. Food and Drug Administration
HTML = hyper text markup language
PRR = proportional reporting ratio
SQL = structured query language