

Research Objectives and Aims

Our project's objective is to develop and evaluate automated methods for detecting unknown adverse drug reactions using heterogeneous data, knowledge-based and statistical-based methods. We initially used comprehensive clinical information, including textual information, in the Electronic Health Records (EHRs) of patients at two institutions combined with the FDA's adverse event reporting system (FAERS), and developed methods to combine the heterogeneous data. However, the use of clinical data from the EHR of one or two institutions does not provide enough power to detect rare adverse events, particularly for medications that are not frequently prescribed. Thus, we would like to use the five large health care datasets (OMOP CDM Version 4.0 consisting of CCAE, MDCD, MDCR, MSLR, and GE) maintained by the IMEDS Research Laboratory. This project aims to develop statistical methods for integrating data from the claims data base along with data from FAERS and from comprehensive data from the EHR. We hypothesize that the addition of the five large health care datasets will complement our data sets and will contribute to development of improved automated methods for detecting novel adverse event signals.

Scope/Proposed Approach

We will continue to use an adaptation of our current methods for integrating heterogeneous data. The approach for using EHR and patient claims data will be based on the following four steps:

- 1) Determine the odds ratios (e.g. the ratio of the odds of a specific AE occurring in the group of patients taking a specific drug versus the group not taking the drugs) for each of a selected set of specific adverse events for each of the health care sites. The different sites will include the EHRs of two healthcare institutions and the five claims datasets.
- 2) Estimate the marginal drug-adverse event associations and stratify them by site using the Mantel Haenszel test to test whether a common odds ratio holds for each of the sites. If the common odds ratio is accepted, we will re-estimate the association assuming a common odds ratio for the common sites; otherwise we will estimate the odds ratio separately for the sites that are not common.
- 3) Identify potential confounders. The development of an AE is generally caused by an underlying condition of the patient. If such condition or co-morbidity leads to the prescription of a certain drug, it could falsely inflate that AE-drug association. In this step, potential confounders are identified using two sets of stratified logistic regressions in conjunction with automatic variable selections using the group LASSO or SCAD penalty. The approach to detecting and handling confounding is described in more detail in Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. [J Am Med Inform Assoc.](#) 2014 Mar-Apr;21(2):308-14.
- 4) Re-estimate the drug-AE associations by conditioning on the identified confounders determined in step 3. The same analysis approach will apply to all the drug-AE pairs.

We will address the multiple comparison issues by using the False Discovery Rate method of Benjamini and Hochberg which is recommended when a large number of comparisons are expected and the traditional Bonferonni approach becomes too conservative.

The approach for combining the patient data and FAERS data will explore use of logistic regression in a manner similar to the one described in the 4 steps above, which was designed to combine patient data from two different sites. Following the above steps, we will quantify the drug-cause-AE association by the odds ratios estimated from applying generalized logistic regressions to the patient data. The patient data consist of health records of millions of patients, and, hence the patients in the EHR data can be viewed as a random sample from the general population. Consequently, the estimated odds ratio represents the relative risk with and without the drug for any patient in the general population.

Impact

Development of new automated methods for detection of novel adverse events based on leveraging EHR data, claims data, and FAERS data could lead to discovery of unknown adverse drug reactions and to improved patient safety. The methods will be published and disseminated to the public.

Experience

We have over 20 years of experience in working with EHR data, with natural language processing to obtain structured coded data from textual patient records, and with design of data bases to store patient data abstracted from text. Our methods have been shown to improve patient care and to be effective for real clinical applications. We also have experience with knowledge-based methods as well as statistical methods.

Names and Titles of individuals requesting access

Carol Friedman, Professor and Graduate Program Director, Dept of Biomedical Informatics, Columbia University

George Hripcsak, Professor and Chair, Dept of Biomedical Informatics, Columbia University

Ying Wei, Associate Professor, Dept of Biostatistics

Ying Li, PhD student, Dept of Biomedical Informatics, Columbia University

Lyudmila Shagina Ena, Senior Programming Analyst, Dept of Biomedical Informatics, Columbia University

Evelyn Rustia, MD, postdoctoral research fellow, Columbia University

Sivan Kinberg, MD, postdoctoral research fellow, Columbia University

Timeline

This project will extend over the next 2 years. Access is needed as soon as possible. The project is anticipated to end in June 30, 2017.