Exploring the Use of ClinicalTrials.gov Trial Results Data for Pharmacovigilance

Vojtech Huser, Olivier Bodenreider
Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD

Introduction

Pharmacovigilance aims to monitor drug safety using sources such as spontaneous reporting systems, biomedical literature or electronic health record data. Clinical trials represent a source of drug-event pairs data complementary to these sources for signal detection in pharmacovigilance platforms. The advantage of CTG over other pharmacovigilance sources is the large number of negative drug-event pairs (explicit evidence that a given drug is not causing a particular adverse event; count of 0 is reported in the deposited results). With 208,959 trials registered and 20,025 trial results, ClinicalTrials.gov (CTG) is the largest repository of trial summary results (with more than 3700 new trial results deposited per year). CTG trial registration data provide information about trial type, sponsor, arms and interventions. CTG results data further provide trial participant counts, baseline characteristics, outcome measures and, most importantly for our study, significant adverse events. Adverse events (AE) are recorded separately by trial arm (or trial group). Although some data submitted to CTG are structured, such as number of arms or intervention type (e.g., drug vs. procedure), many elements are collected as free text (e.g., the drug(s) used in the trial). This preliminary investigation explores the selection of clinical trials of interest for pharmacovigilance and the feasibility of extracting drug concepts from CTG trial registration data.

Methods

To investigate the proportion of drug trials that can be directly used for pharmacovigilance (without additional manual curation), we analyzed results of interventional trials with drug interventions. We used CTG's tabular data format and the structured element *intervention_type*. To identify drug names in the CTG *intervention_name* field, we mapped them to RxNorm using increasingly aggressive techniques, namely using the findRxcuiById (exact/normalized match) and getApproximateMatch functions of the RxNorm API.

Preliminary Results

Feasibility counts: As of February 22, 2016, the tabular CTG data included a total of 14,007 results of interventional trials that had at least one drug intervention (dataset S1; supplemental data are available at github.com/vojtechhuser/CTG). We found that 5,192 trials (37% of our sample) have exactly one trial arm and drugs from such trials can be unambiguously associated with AEs reported in the deposited trial summary results. Additionally, in 2,049 two-arm-trials (14.6% of our sample) we converted their free-text specified placebo arms into formally modelled placebo arms which may allow us to use additional pharmacovigilance methods. Overall, a total of 7241 trials would be amenable to processing for pharmacovigilance purposes.
Drug term detection: We processed 63,817 unique interventions strings (extracted from all CTG interventions of type 'drug'). For 10.4% of those strings, the RxNorm API exact or normalized match function identified RxNorm concepts of type ingredient, clinical drug or branded drug (dataset S2). When no exact or normalized match was found, we proceeded with approximate match and were able to map additional 0.6% of input strings without the need for human review (single RxCUI detected with 100% detection certainty score; dataset S3). For the remaining input strings, approximate match provided multiple inputs (6.1 RxNorm terms on average) with a wide range of scores.

Conclusions

These results indicate that intervention names are usually not simple drug names and would require parsing for extraction of drug names (e.g., with MedEx, a medication text processing tool). Overall, our work indicates that a significant portion of the CTG result database can generate some drug-adverse event pairs that can be used for pharmacovigilance purposes. In our experience, the main obstacle to leveraging CTG for pharmacovigilance is the difficulty in unambiguously associating information from registration (intervention drug + trial arms) with information from result summaries (AEs), because of the absence of an explicit link between trial arms in these two resources.

Acknowledgments: This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.