Incidence-based hierarchical clustering analysis of adverse drug events in the FAERS database

Lei Wang, PhD\textsuperscript{1}, Lang Li, PhD\textsuperscript{1}

\textsuperscript{1} Department of Biomedical informatics, The Ohio State University College of Medicine, Columbus, Ohio

Introduction

The FDA Adverse Event Reporting System\textsuperscript{1} (FAERS) can monitor unexpected drug outcomes over a large population of patients in the form of adverse drug event (ADE) reports. ADEs and medication errors are coded using the Medical Dictionary for Regulatory Activities\textsuperscript{2} (MedDRA) terminology in the FAERS database. MedDRA has a 5-level, organ-based hierarchical structure, which provides the functions of both coding adverse effects and grouping related terms for database searches.

Although MedDRA improves the accuracy of coding ADE, it has some limitations: It does not provide formal definitions of terms; The multiaxial linkages of terms are incomplete, etc. Clustering ADRs from the FAERS can show us an unbiased characteristics of ADEs and their similarities. In this paper, we utilized a hierarchical clustering method to identify the consistent or evolving ADE patterns from 2004 to 2018 by System Organ Classes (SOCs) and High Level Group Terms (HLGTs) in the MedDRA.

Methods

FAERS data processing: The event reports submitted from 2004 to the third quarter of 2018 were downloaded from the FFAERS. To eliminate the data redundancy caused by multiple case versions in FAERS system, duplicated cases were removed based on the following data fields: case id, case initial/follow-up code, event date, age, sex, reporter country, drug names, and reaction/outcomes preferred terms). If cases match all these fields, only the latest case report was kept. In addition, if case reports contained the same symptoms both in ADEs and drug indications, those ADEs were removed. After the data processing, there were 9,237,447 ADE reports, which contained a total of distinctive 19,904 ADEs. ADEs in the FAERS database were annotated using the MedDRA.

Unsupervised hierarchical clustering: In order to find underlying clusters of ADEs in a patient incidence profile similarity, we performed unsupervised hierarchical clustering on the constructed ADE-case matrix. Briefly, a set of similarities were generated from the matrices using Jaccard distance\textsuperscript{4}, and each element (SOC or HLGT) was assigned to a cluster; iteratively. Clustering was performed using Ward’s linkage\textsuperscript{5} with the \texttt{hclust} function in the R \texttt{stats} package.

Results

The cluster dendrogram reproduced in Figure 1. As demonstrated in Figure 1(A), the incidence of ADE is generally grouped into three clusters on the SOC level. We found that some severe and chronic ADEs have similar incidence patterns, e.g., Cardiac disorders (Card), Vascular disorders (Vasc), Metabolism and nutrition disorders (Metab) and Renal and urinary disorders (Renal). Further, we conduct HLGT level based hierarchical clustering analysis for these four SOCs (Figure 1B). The cluster dendrogram indicates that HLGTs are also grouped into three clusters. Particularly, Arteriosclerosis, stenosis, vascular insufficiency and necrosis and Coronary artery disorders share the same incidence patterns (distance is 0.3714).
Discussion

Although MedDRA is widely used in ADE spontaneous reporting database, it has some limitations. We perform a hierarchical clustering analysis in the FAERS database using SOCs and HLGTs. The results demonstrate that in HLGTs, Arteriosclerosis stenosis, vascular insufficiency and necrosis (AVIN) and Coronary artery disorders (CAD), share similar ADE patterns. However, these two terms are solely linked to Vascular disorders and Cardiac disorders separately on SOC level in MedDRA, which shows no linkage between them. This indicates that they may have a potential missing linkage.

References