Use of Natural Language Processing Text-Mining to Identify Differences in the OVERDOSAGE Section of Drug Labeling

Ashraf, Adrita, FDA/CDER; Brodsky, Eric, FDA/CDER; Burkhart, Keith, FDA/CDER

Abstract

Background: The OVERDOSAGE section of labeling of different drugs with the same active ingredient typically have similar language. However, sometimes there are differences in the content of the OVERDOSAGE section of labeling for drugs with the same active ingredient (e.g., due to differences in the conditions of use of the drugs). A manual review of the OVERDOSAGE section of labeling for a specific active ingredient to identify differences may be labor intensive. Natural Language Processing (NLP) text mining tools may facilitate the process of identifying similarities and differences in the OVERDOSAGE section of labeling with drugs with the same active ingredient.

Purpose: Perform a test case to evaluate the ability of an automated NLP text mining tool to search a labeling database, identify and extract targeted information from the OVERDOSAGE section of labeling, generate structured output, and analyze for similarities and differences in the wording in the OVERDOSAGE section of labeling for a specific active ingredient.

Methodology: A query was developed using the text-mining platform, Linguamatics. The query searched the OVERDOSAGE section of drug labeling with the same active ingredient identified by its Unique Ingredient Identifier (UNII) using structured product labeling files in DailyMed. Subsequently, the wording in the OVERDOSAGE section of the labeling for this active ingredient was searched and analyzed for similarities and differences. Results: The query retrieved 48 different labeling (Physician Labeling Rule (PLR) format and non-PLR format) labeling for drugs with the active ingredient including 38 fixed combination drug products (containing the active ingredient and at least one additional active ingredient) and 10 single ingredient products. Among 48 labeling, 17% were for prescription drugs approved under New Drug Applications (NDAs) and 83% were for prescription generic drugs approved under Abbreviated New Drug Applications (ANDAs). Of the 8 labeling under NDAs, 6 had different content in the OVERDOSAGE section. Of the 40 labeling under ANDAs (generic drug labeling), there were two sets of labeling with different content in the OVERDOSAGE section.

Conclusions: Natural Language Processing text mining can be used to query labeling in DailyMed and identify the similarities and differences in the OVERDOSAGE section labeling for a specific active ingredient.

Introduction

• Drug overdosage is the leading cause of injury-related deaths in the United States, surpassing deaths from motor vehicle accidents and homicides over the past two decades.1
• According to Food and Drug Administration (FDA) regulations, the OVERDOSAGE section of labeling for human prescription drugs must include: signs, symptoms, laboratory findings, and complications of overdose; drug concentration associated with toxicity; amount of drug associated with overdose and the amount of drug that is likely to be life-threatening; dialyzable information; and recommended overdose treatment.2
• FDA regulations require that the OVERDOSAGE section of labeling is updated when new information becomes available that causes the labeling to be inaccurate, false, or misleading.3
• CDER is working to update the OVERDOSAGE section of labeling for drug classes associated with the most fatalities according to National Poison Data System (NPDS) data4 from the American Association of Poison Control Centers (Fig. 1).
• Natural Language Processing (NLP) text mining can be used to efficiently query labeling in DailyMed and identify the similarities and differences in the OVERDOSAGE section of labeling for a specific active ingredient.

Materials and Methods

• A machine learning based NLP text-mining platform, Linguamatics, was used to query structured product labeling files in the National Institutes of Health’s (NIH) DailyMed labeling database.
• For one active ingredient (i.e., active ingredient X) within the drug class listed above, the query extracted all the labeling in the drug class by using active ingredient X’s Unique Ingredient Identifier (UNII).
• For each identified labeling, the wording in the OVERDOSAGE section of the labeling was extracted (Fig. 2) and then reviewed manually for similarities and differences.

Results and Discussion

• The query generated a structured output tabulating the drug name(s), dosage form(s), strength(s), marketing category (e.g., NDA, ANDA), application number(s), and date that the OVERDOSAGE section was updated. (Fig. 3)
• The query retrieved 48 different labeling (Physician Labeling Rule (PLR) format and non-PLR format labeling) containing active ingredient X (associated with 86 application numbers). Of the 48 labeling, 38 were fixed combination drug products (active ingredient X and Y) and 10 were single ingredient products (active ingredient X).

• Among these 48 labeling, 17% were for prescription drugs approved under New Drug Applications (NDAs) and 83% were for prescription generic drugs approved under Abbreviated New Drug Applications (ANDAs).
• Of the 8 labeling under NDAs, 6 had some differences in content in the OVERDOSAGE section due to different conditions of use of the drugs.
• Of the 40 labeling under ANDA (generic drug labeling) one set of labeling (n=31 labeling) had some differences in content in the OVERDOSAGE section than another set of labeling (n=9 labeling) (the two sets of labeling were for two different dosage forms).
• Additional query and algorithm development may further reduce the manual component of these analyses.

Conclusion

Natural Language Processing text mining can be used to efficiently query labeling in DailyMed and identify the similarities and differences in the content of the OVERDOSAGE section labeling for a specific active ingredient.

References

1. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). Underlying cause of death 1999–2016 on CDC WONDER online database, released December, 2017. Data are from the Multiple Cause of Death Files, 1999–2016, as compiled from data provided by the 57 Vital Statistics Systems under the authority of the National Center for Health Statistics Cooperation Program.

Disclaimer: This poster reflects the views of the authors and should not be construed to represent the FDA’s views or policies.

Figure 1. Drug Classes Associated with the Highest Number of Overdose Fatalities (reported by the NPDS)

Figure 2. Linguamatics Query to Extract Wording in the OVERDOSAGE Section for Labeling with Active Ingredient X.

Figure 3. Linguamatics Output – List of Labeling with Different Wording in the OVERDOSAGE Section for Active Ingredient X.