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

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





Abstract

Background: The OVERDOSAGE section of prescription drug labeling with the same active ingredient may be different (e.g., due to differences in the conditions of use of the drugs). A search of the OVERDOSAGE section of labeling for drugs with the same active ingredient in order to identify content differences may be laborious. Natural Language Processing (NLP) text-mining tools may facilitate the process of identifying these differences.

Purpose: This will evaluate the ability of an automated text-mining tool to search a labeling database, identify and extract targeted information from the OVERDOSAGE section for drugs with the same active ingredient and generate structured output to analyze for differences.

Methodology: Using a text-mining platform, Linguamatics, a query was developed that 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 text was analyzed for differences

Results: The query retrieved 48 prescription drug labeling [Physician Labeling Rule (PLR) format and non-PLR format labeling] with the same active ingredient. Among the 48 labeling, 17% (8) were drugs approved under a New Drug Application(s) (NDA); while 83% (40) were generic drugs approved under an Abbreviated New Drug Application(s) (ANDA). Text-mining identified 6 different OVERDOSAGE sections among the 8 NDA labeling and 2 different OVERDOSAGE sections among the 40 ANDA labeling.

Conclusions: Natural Language Processing text-mining can be used to query labeling in DailyMed to identify the differences in the OVERDOSAGE section labeling for drugs with the same 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.^{1,2}
- 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 overdosage; drug concentration associated with toxicity; amount of drug associated with overdosage and the amount of drug that is likely to be life-threatening; dialyzable information; and recommended overdosage treatment.³
- 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.⁴
- 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) data⁵ 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.

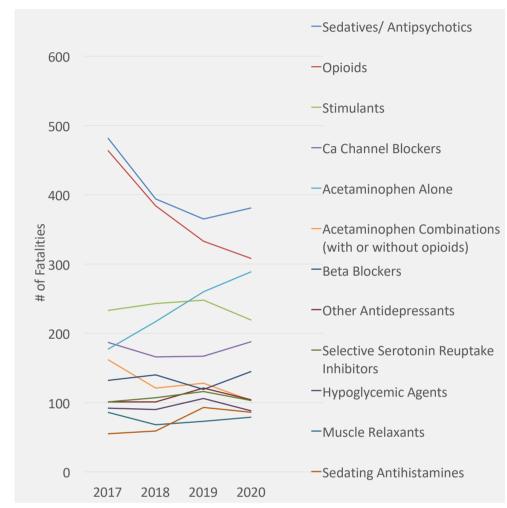


Figure 1. Drug Classes Associated with the Highest Number of Overdosage Fatalities (reported by the NPDS)

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).

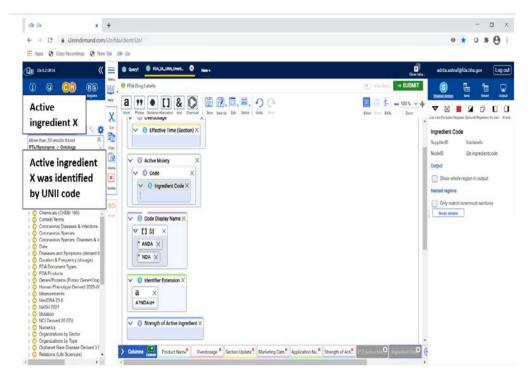


Figure 2. Linguamatics Query to Extract Wording in the OVERDOSAGE Section for Labeling with Active Ingredient X.

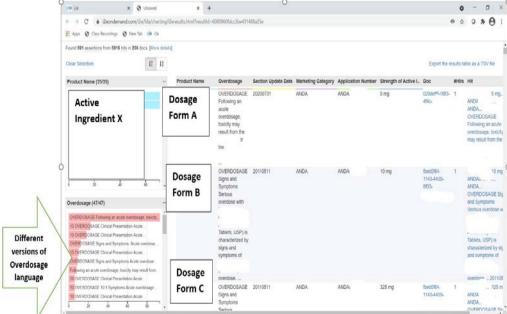
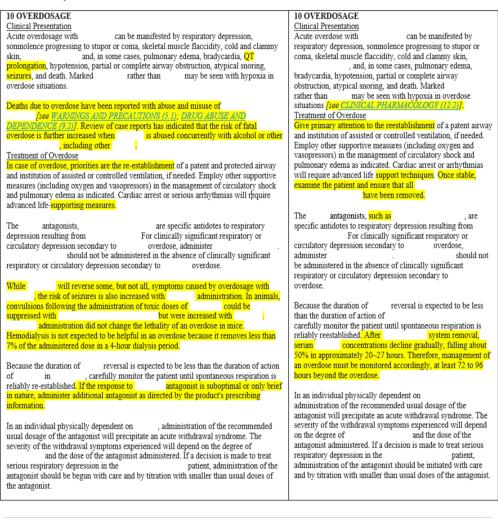


Figure 3. Linguamatics Output – List of Labeling with Different Wording in the OVERDOSAGE Section for Active Ingredient X.

- 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 ANDAs (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.

Table 1. Representative Example of the Similarities and Differences (highlighted in yellow) Between Two OVERDOSAGE Sections of Labeling for Drugs with Active Ingredient X (drug names were redacted).



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). Compressed mortality file 1979–1998. CDC WONDER on-line database, compiled from compressed mortality file CMF 1968–1988, Series 20, No. 2A, 2000 and CMF 1989–1998, Series 20, No. 2E, 2003.

2.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 jurisdictions through the Vital Statistics Cooperative Program.

3. Code of Federal Regulations - 21 CFR 201.57(c)(11): https://www.ecfr.gov/cgi-bin/text-

idx?SID=50438b5cc7161b2bd5102ff1abb9fb4f&mc=true&node=se21.4.201_157&rgn=div8_4. Code of Federal Regulations - 21 CFR 201.56(a)(2): https://www.ecfr.gov/cgibin/retrieveECFR?gp=&SID=f06bb31d261a75f91e9fb5264da0e2fe&mc=true&n=sp21.4.201.b&r=SUBPART&tv=HTML#se21.4.201_156

5. National Poison Data System Annual Reports - Table 18. Categories Associated with Largest Number of Fatalities: https://www.npds.us/Reports

Acknowledgement: Ryan Hulscher (Linguamatics).

<u>Disclaimer:</u> This poster reflects the views of the authors and should not be construed to represent the FDA's views or policies.