TITLE PAGE

Type of Paper:

Commentary

Title:

The Irony of MedWatch and the FAERS Database: An Assessment of Data Input Errors and Potential Consequences

Authors' full names:

Michael A. Veronin, B.S.Pharm., M.S., Ph.D.* * Corresponding author:
Associate Professor
The University of Texas at Tyler
Ben and Maytee Fisch College of Pharmacy
Department of Pharmaceutical Sciences
3900 University Blvd.
Tyler, Texas 75799
Tel: (903) 566-6148
Fax: (903) 565-5598
E-mail: mveronin@uttyler.edu

Robert P. Schumaker, Ph.D.
Professor
The University of Texas at Tyler
Soules College of Business
Department of Computer Science
3900 University Blvd.
Tyler, Texas 75799
Tel: 903-565-5578
Fax: (903) 565-5598
E-mail: rschumaker@uttyler.edu

Rohit Dixit, M.S.
The University of Texas at Tyler
Soules College of Business
Department of Computer Science
3900 University Blvd.
Tyler, Texas 75799
Tel: (903) 566-7403
E-mail: rohitdixit188@live.in

Introduction

> **Irony** [ahy-ruh-nee]. Noun.
> "An outcome of events contrary to what was, or might have been, expected."
> Source: http://www.dictionary.com.

In healthcare, accurate documentation and records management is important for clinical care and research as well as ensuring patient safety. It is crucial for documentation errors to be corrected to ensure that medical records are interpreted correctly.

One of the most common problems with data management occurs during the data input process.[1] It is reasonable to assume that even a seemingly insignificant typographical error can cause short and long-term problems, which may lead to inaccurate records, misinformation, and disorganization. This is particularly relevant in instances of manual data entry.

This report presents a brief overview of input errors in one of the nation's largest database for medication error reporting, the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). The focus of this narrative is on errors and inconsistencies in reporting of drug names in this database, initiated through data input of the MedWatch reporting system.

Background

In healthcare, when an error occurs with the use of a medication, (i.e., "medication error"), an official process exists for reporting of the error. As asserted by Elden and Ismail (2016), "Error detection through an active management and effective reporting system discloses medication errors and encourages safe practices."[2]

The FDA Adverse Event Reporting System, or FAERS, is one of the largest government databases in the country. Maintained by the U.S. Food and Drug Administration (FDA), FAERS includes reports on adverse event and medication error reports that have been submitted to the FDA.[3]

FAERS is known as an essential source of data in monitoring drug effects in order to identify and evaluate previously unreported adverse events.[4] Organized in tables, the FAERS database is structured as a computerized, relational database consisting of seven file "packages".[5] Items of information contained within the file packages, and linkable to other files within the database include: patient demographics, drug names, indications, outcomes, reactions, report sources, and therapies. The DRUG file contains the DrugName data set, consisting of the names of the reported medications, which are suspect for Adverse Drug Events (ADEs).

Founded in 1993, the MedWatch system includes records created in the FAERS database from reports of adverse events and medication errors by healthcare professionals, consumers, and drug manufacturers.[6]

MedWatch reporting involves completion of a print or online "fillable" form. Individual reports of medication errors, product quality problems, and adverse drug reactions are either sent directly using "Medwatch form 3500" to the FDA, or indirectly via the product's manufacturer, who is required by law to submit a "MedWatch form 3500A" upon receipt of adverse event information from a health professional or consumer.[7] The text fields in the forms are fillable via computer using the free Adobe Acrobat Reader. Users enter the adverse event report data, including patient information, a description of the adverse event, and the suspect product. "Suspect products" may include drug name and dosage, outcome attributed to an adverse event, and other related information as described on the form.

Methods

Data Source

Data analysis was conducted at the Data Analytics Laboratory in the Computer Science Department at our academic institution. As part of ongoing research, drug name records in the

DrugName file in the FAERS database were obtained, which yielded a corpus of over 32 million records. Drug names were "cleaned" (i.e., rectified) and standardized by a unique, multi-step process.

During the cleaning and standardization process, variations in drug names were identified as potential input errors. Drug names encountered included generic, brand, and combination drug product names. Infrequently, a drug's scientific or chemical, or investigational drug name was observed, presumably because some health care institutions, such as teaching hospitals that utilize MedWatch, serve as clinical research sites.

The process involved a combination of automated and manual methods: recognition of potential drug name aberrations, check against established drug name references, and transformation to standardized format. Subsequently, the identified variant drug names were categorized to distinguish the types of erroneous input. Details of the drug name standardization process are provided as Supporting Online Material in the References section.

Results

At least 19 types of errors emerged from analysis of the FAERS drug name data. Information on the type of drug name variation with examples, and potential correction measures is presented in Table 1.

Discussion

Broadly speaking, an "error" is defined as "a deviation from accuracy or correctness; a mistake, as in action or speech".[8] The FAERS database was created for adverse drug event (ADE) reporting, which includes medication errors. ***Ironically, the federal government's premier system for reporting drug errors includes several types of errors in the reporting itself.***

As often the case, errors are not without consequences, which may include impact on data integrity, research, and patient safety.

Consequences for Maintenance of Data Integrity

The term "data integrity" refers to the accuracy and consistency of data.[9] In addition to input errors as described here, data integrity also can be compromised by other means, such as when data is transmitted from one computer to another, by software "bugs" or viruses, hardware malfunctions (i.e., disk "crashes,") and natural disasters (i.e., fires, floods). There are many ways to minimize these threats to data, most notably designing user interfaces that prevent the input of invalid or incorrect data, and using error detection and correction software.[10] Our findings reveal that neither of these approaches is being used in the MedWatch reporting system and FAERS database maintenance.

The drug name data records contained null values, ambiguous or nonspecific terms, upper and lowercase letters or both, leading and trailing whitespace, new-line and tab characters, leading numbers, special characters, drug name combinations with no delineation of entities, abbreviations, misspellings.

These examples represent a wide range of data inconsistencies, and coupled with the sheer volume of data in one of the U.S government's largest healthcare databases, FAERS, it is clear that data integrity can be one of the biggest challenges in data management.

A large number of tools of varying functionality is available to support data cleaning tasks, but often a significant portion of the cleaning and transformation work has to be done manually or by low-level programs that are difficult to write and maintain, which adds to the complexity.[11]

It is little wonder that a recent survey of data scientists ascertained that data preparation accounts for about 80% of the work of data scientists, and cleaning data is the least enjoyable and most time-consuming data science task.[12]

Consequences for Research

Data is the lifeblood of research, so that the correctness of data at all stages of a research project is vital to avoid wrong conclusions. For instance, duplicated or missing information will produce incorrect or misleading statistics, or in the vernacular of computer scientists, "garbage in, garbage out."

The FAERS database is known widely as an essential source of data in the monitoring of the effects of drugs in order to identify and evaluate previously unreported adverse events. For research on safety evaluations to be valid, they must be based on FAERS data that is accurate, and complete; essentially "cleaned" in the best possible manner to remove aberrations such as those presented in Table 1. In the case of a drug, if it is underrepresented due to misrepresentation in the database, it can affect the validity of the research on identifying adverse drug events.

Consequences for Patient Safety

Input errors in documents related to healthcare delivery have significant implications on patient safety. The goal of reporting systems in adverse drug events and medication errors is to reduce the likelihood of harm related to medications.

The Food and Drug Administration credits the MedWatch system with improving awareness, and expediting early detection, of drug and device risks and in illuminating the adoption of medical treatments.[13] Data input errors resulting in inaccurate adverse event reporting may lead drug safety professionals to draw incorrect conclusions, manufacturers may be wrongly forced to

suspend and withdraw medications and interventions, health professionals may mistakenly alter their clinical practices, and patients may be denied safe and effective treatments.

Conclusion

An article by Gellert (2016) in *British Medical Journal* asserts that "U.S. healthcare is the last major American industry to have its information infrastructure become electronic and digital…relatively new technologies still remain immature and in need of substantial improvements in usability, functionality and interoperability."[14]

The FDA Adverse Event Reporting System (FAERS) collects Adverse Drug Event data from the U.S. and Europe through the MedWatch reporting system, and offers a public access to researchers and consumers.[3,4] As often the case with large datasets, such as drug name data maintained in the FAERS database, it is necessary to detect incomplete, inaccurate, or inappropriate data and implement corrective measures where needed. The consequences from erroneous data input – in this case, drug names - can have an impact on data integrity, research, and patient safety. Insight into the nature of input data – and corresponding problems that occur - has the potential to prove beneficial as researchers move towards making the FAERS database usable at the clinical practice level.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Lai KH, Topaz M, Goss FR, Zhou L. Automated misspelling detection and correction in clinical free-text records. *J Biomed Inform*. 2015 Jun;55:188-95. doi: 10.1016/j.jbi.2015.04.008. Epub 2015 Apr 24.

2. Elden N, Ismai A. The Importance of Medication Errors Reporting in Improving the Quality of Clinical Care Services. *Glob J Health Sci.* 2016 Aug; 8(8): 243–251. doi: 10.5539/gjhs.v8n8p243.

3. U.S. Food and Drug Administration. "Questions and Answers on FDA's Adverse Event Reporting System (FAERS). What is FAERS?" Available from: https://www.fda.gov/drugs/surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers. Accessed April 25, 2020.

4. Pratt LA., Danese PN. More eyeballs on AERS. Nat Biotechnol 2009; 27 (7):601-602. doi: 10.1038/nbt0709-601.

5. Poluzzi E, Raschi E, Piccinni C, de Ponti F. Data mining techniques in pharmacovigilance: analysis of the publicly accessible FDA Adverse Event Reporting System (AERS). In: Karahoca A (Ed): Data mining applications in engineering and medicine. Croatia: InTech; 2012:267–301. doi: 10.5772/50095.

6. Kessler DA. Introducing MEDWatch. A new approach to reporting medication and device adverse effects and product problems. *JAMA*. 1993 Jun 2;269(21):2765-8. doi:10.1001/jama.1993.03500210065033.

7. U.S. Food and Drug Administration. MedWatch: The FDA Safety Information and Adverse Event Reporting Program. Available from: http://www.fda.gov/Safety/MedWatch/default.htm. Accessed April 25, 2020.

8. “Error” Definition 1. Dictionary.com. 2019. Available from: http://www.dictionary.com/browse/error?s=t. Accessed April 25, 2020.

9. Database.Guide. What is data integrity? Available from: https://database.guide/what-is-data-integrity/. Published May 28, 2016. Accessed April 25, 2020.

10. Beal V. Webopedia. Data integrity. 2019. Available from: https://www.webopedia.com/TERM/D/data_integrity.html. Accessed April 25, 2020.

11. Rahm, E., Do, H. D. (2000). Data cleaning: problems and current approaches. *IEEE Data Eng Bull.* 2000; 23(4):3-13.

12. L.V. Least enjoyable and most time-consuming data science tasks. *Data Science Central*. Available from: https://www.datasciencecentral.com/profiles/blogs/bubu. Published April 5, 2016. Accessed April 25, 2020.

13. Getz KA, Stergiopoulos S, Kaitin KI. Evaluating the completeness and accuracy of MedWatch data. *Am J Ther*. 2014 Nov-Dec;21(6):442-6. doi: 10.1097/MJT.0b013e318262316f.

14. Gellert, G.A. (2016). Rapid Response. EHRs, CPOE and Digital documentation: an effective component of the U.S. national response to the epidemic of medical error related deaths. *BMJ.* 2016;353:i2139. Available from: https://www.bmj.com/content/353/bmj.i2139/rr-17. Accessed April 25, 2020.

Supporting Online Material

https://drive.google.com/open?id=1iPI-UJnwWFB6aUIc39Bwml9qUyprPQxP

Table 1. Drug Name Input Errors in MedWatch and the FAERS Database with Corrective Action

Drug Name Variation	Example	Potential Corrective Action*	Comments
Null value	"NULL"	Entry needed in input field	The term "null" is returned when input field is blank
Nonspecific/ ambiguous drug description	"Pain med", "antibiotic"	Specific drug name needed in input field	User inputs general drug category for drug name
Only brand name represented without generic name	"Humira"	Brand name converted to generic	Brand names may vary; consistent presentation needed
Inconsistent Upper/Lower case drug name description	"Aspirin", "aspirin", "ASPIRIN"	Generic names converted to upper or lower case; brand names capitalized	Brand names are usually capitalized while generic names are not
Leading/Trailing whitespace	"␣aspirin", "aspirin␣", "␣aspirin␣"	Character spaces removed manually or with text editor tool	Does not correspond to a visible mark, but occupies an area in drug name
Keyboard new line/carriage return inserted	"<CR >aspirin", "aspirin <CR >"	Spaces removed manually or with text editor tool	Does not correspond to a visible mark, but occupies an area in drug name
Tab character inserted	"»acetaminophen"	Spaces removed manually or with text editor tool	Does not correspond to a visible mark, but occupies an area in

			drug name
Leading numbers	“123Humira”	Numbers removed manually or with text editor tool	Correction requires knowledge of proper drug name
Special characters inserted	“@Tylenol”	Characters removed manually or with text editor tool	Likely due to typographical error; Correction requires knowledge of proper drug name
Drug name combinations as single entity	“lisinopril hydrochlorothiazide”	Single entries created with names separated by “and”	Correction requires knowledge of proper drug combination name
Abbreviations	“Vit B12”	Names expanded to complete description (i.e. “Vitamin B12”)	Unapproved abbreviations not recommended
Extra identifier	Humira (Abbott)	Extra name removed manually or with text editor tool	Manufacturer included in drug name
Omitted letters	“ateplase”, “oxycodon”	Correct name identified (i.e., “alteplase”, “oxycodone”)	Likely due to typographical error or knowledge deficit
Extra letters	“dicloflenac” “chlotrimazole”	Correct name identified (i.e., “diclofenac”, “clotrimazole”)	Likely due to typographical error or knowledge deficit
Hyphens inserted	“hydro-chlorothiazide”, hydroxy-chloroquine	Correct name identified (i.e., “hydrochlorothiazide”,	Incorrect application of rule: “insert a hyphen between a

		“hydroxychloroquine”)	prefix and a proper noun (name)”
Transposition of letters	“nuetrexin”	Correct name identified (i.e., “neutrexin”)	Often occurs with reversed order of double vowels
Wrong letter	“acetaminofen” “filgrastin”	Correct name identified (i.e., “acetaminophen”, “filgrastim”)	Likely due to knowledge deficit
Incorrectly Repeated Consonants	“melphallan”	Correct name identified (i.e., “melphalan”)	Likely due to knowledge deficit
Wrong Vowel	“netilmycin”, “amitriptiline”	Correct name identified (i.e., “netilmicin”, “amitriptyline”)	Likely due to knowledge deficit

*Text editor tool is not available in the online MedWatch form