Study of the Drug-related Adverse Events with the Help of Electronic Health Records and Natural Language Processing

Sarah Allabun[1]*, Ben Othman Soufiene[2]
Department of Medical Education-College of Medicine, Princess Nourah bint Abdulrahman University,
P.O.Box 84428, Riyadh 11671, Saudi Arabia
PRINCE Laboratory Research-ISITcom-Hammam Sousse, University of Sousse, Tunisia

Abstract*—Surveillance of pharmacovigilance, also known as drug safety surveillance, involves the monitoring and evaluation of drug-related adverse events or side effects to ensure the safe and effective use of medications. Pharmacovigilance is an essential component of healthcare systems worldwide and plays a crucial role in identifying and managing drug safety concerns. Natural language processing (NLP) can play a crucial role in surveillance activities within pharmacovigilance by analyzing and extracting information from various sources, such as clinical trial reports, electronic health records, social media, and scientific literature. It is important to note that while NLP can be a powerful tool in pharmacovigilance surveillance, it should always be used in conjunction with human expertise. NLP algorithms can assist in the identification and extraction of relevant information, but the final assessment and decision-making should involve the knowledge and judgment of trained pharmacovigilance professionals. In this paper, we intend to train and test our models using the dataset from the Medication, Indication, and Adverse Drug Events challenge. This dataset will include patient notes as well as entity categories such as Medication, Indication, and ADE, as well as various sorts of relationships between these entities. Because ADE-related information extraction is a two-stage process, the outcome of the second step (i.e., relation extraction) will be utilized to compare all models. The analysis of drug-related adverse events using electronic health records and automated approaches can considerably increase the effectiveness of ADE-related information extraction, although this depends on the methodology, data, and other aspects. Our findings can help with ADE detection and NLP research.

Keywords—Natural language processing; surveillance of pharmacovigilance; drug-related adverse

I. Introduction

Adverse drug events are caused by medications, and these forms of injuries are referred to as adverse drug events (ADEs). A few years ago, data mining techniques for identifying adverse drug events (ADEs) by analyzing information were discovered. These data mining tools examine complex data extracted from huge electronic medical databases [1, 2]. The International Standard Organization (ISO) defines electronic health database (EHR) as a repository of patient-related databases in digital form, securely stored and shared, and accessible by various authorized healthcare users. Clinical information such as frequency of drug use, adverse effects, and so on can be found in electronic databases [3]. Randomized controlled clinical trials (RCTs) are considered the gold standard for investigating the pros and cons of drugs; however, due to limitations such as shorter duration and restricted inclusion criteria, the development of data mining databases and algorithms for Pharmacovigilance tasks has resulted [4]. Since 1960, the most widely available type of medical database for adverse events has been the spontaneous reporting system (SRS) database. The notable SRS databases are the Adverse Event Reporting System (ARRS), the Medicines and Healthcare Products Agency's (MHRA) Yellow card scheme, the European Agency for the Evaluation of Medicinal Products (EMEA), and the World Health Organization. The Adverse Event Reporting System (ARRS) and the General Practice Research Database (GPRD) are still operational. ARRS is a well-known adverse event database that supports the FDA's safety program for all approved pharmaceuticals, while the GPRD database (Pharmacoepidemiology database) is a significant database of medical records [5].

The significance of our work is that several phase clinical trials assess the adverse effects of each licensed drug, whereas clinical studies typically target only one drug. The specific effects of multiple-drug administration become harder to analyze. People, on the other hand, take many medications, which creates a chasm between research trials and actual drug use by patients. As a result, information about Adverse Drug Reactions (ADEs) is critical for ensuring patient safety [5]. Clinical information cannot be retrieved from biomedicine literature or narrative clinical reports; consequently, natural language processing (NLP) has been developed expressly for this purpose, which identifies, extracts, and encodes information from biomedicine literature and clinical narratives [4].

The practice of collecting useful information from vast amounts of unstructured text using computational algorithms is known as text mining [6,7]. In the context of pharmacovigilance, "meaningful information" is information that can aid in the detection and assessment of adverse drug events (ADEs). Because text mining provides a technique for converting free text into computable knowledge, it is developing as a method for exploring, analyzing, querying, and managing unused drug safety information. Pharmacovigilance is now based on the analysis of clinical trials and spontaneous reports, as well as a review of biological literature to some extent. Domain experts often do the study on a case-by-case basis. Statistical approaches have recently been included into standard Pharmacovigilance practice, and these are applied to spontaneous reports [8, 9] and clinical trials [10] to further discover ADE signals. Nonetheless, well-known limitations [11, 12] inherent in the type and range of data sources used in routine Pharmacovigilance, as well as rising public concern about medication safety, have sparked a slew of global research and legislative initiatives [13, 14] aimed at enhancing Pharmacovigilance. It is well acknowledged that development in pharmacovigilance is dependent on a comprehensive methodology that analyses ADE-related data from a diverse range of potentially complementary data sources. With the passage of the Food and Drug Administration Amendments Act (FDAAA) of 2007 [15], Pharmacovigilance research has centered on the expanding secondary use of electronic health records (EHRs) [16]. Other sources, such as biological literature, product labels, social media content, and logs of information seeking activities on the Web, have been explored in recent years to assist holistic Pharmacovigilance. Each source offers a distinct point of view, and each has its own set of advantages and disadvantages.

EHRs contain the promise of active surveillance, the ability to quantify the incidence or risk of ADEs, the ability to identify at-risk patients, and the possibility to deliver more accurate and earlier ADE identification. Unlike the current manual approach, it is conceivable to utilise the biomedical literature computationally for a variety of Pharmacovigilance goals, including signal detection [17]. Product labels offer a wealth of information spanning from adverse medication responses to drug efficacy, risk management, contraindications, drug interactions, and many other topics. Several attempts have evolved to computationally extract information from product labels in order to build a database of known ADEs [18]. The generated knowledgebase can be utilized for additional ADE assessment, determining benchmarks for signal identification, prioritizing and filtering ADEs under research, and detecting class effects. Social media, for example, patients' experiences with pharmaceuticals that are explicitly shared through online health forums and social networks, as well as implicit health information contained in major search engine search logs, are among the potential data sources. Text mining combines a wide range of statistical, machine learning, and linguistic approaches related to natural language processing to address the issues given by unstructured text (NLP). It is helpful to think of text mining as a process that employs tools, methods, and heuristics created by those who study natural language processing. Text-mining workflows can employ varying degrees of sophistication in NLP approaches, depending on the use case. As a result, unlike traditional NLP, which employs sophisticated language models and computationally intensive syntactic and semantic analyses to extract meaning from text, text mining favors the use of simpler but less costly approaches that scale to large data sets.

Typically, a text mining process begins with multiple pipelined NLP subtasks that structure the text in preparation for the statistical analysis or pattern identification phase. A set of foundational low-level syntactic activities and a set of high-level tasks that build on the low-level tasks and entail semantic processing are among the subtasks.

The primary goal of this study is to use natural language processing technology (NLP) to extract potential adverse events from the notes section of electronic health records, and then to use traditional bio statistical approaches to detect associations with specific medications that patients are taking.

II. Methods and Materials

A. Population and Study Sample

Patient data from Columbia University Hospital, which contains over free text documents such as correspondence, discharge letters, and events, and are growing at a rate of new documents each month, will be used for research purposes. The Clinical Record Interactive Search System (CRIS) [19], a de-identified version of the EHR, will be implemented to create a research resource, and we are also planning to enhance it with language processing tools to extract information from the vast amount of free text format data stored within our database. The clinical dataset Columbia University EHR Structured and Unstructured ADE corpus, which contains information for all patients at Columbia University Hospital, will be used in this investigation. Data will be extracted from the EHR in Columbia University hospital. The data is gathered from 10th June 2019 to 16th August 2019.

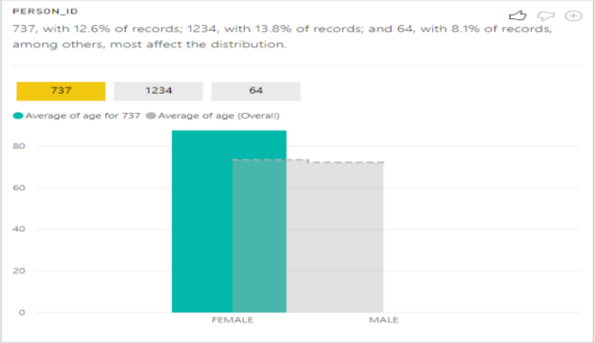
B. Study Significance

The major goal of medication safety regulators and researchers is to discover and monitor ADEs that have the potential to cause public harm [20]. Many ADEs are discovered after a medicine has been commercialized, when it is taken by a broader and more diverse population than in clinical trials [21]. Because ADEs detected after a drug has been widely used can be a significant cause of morbidity and mortality, good post-marketing drug safety surveillance is crucial for public health protection. Only after a new drug's efficacy and safety have been demonstrated in a series of clinical studies is it granted regulatory approval. Randomized, controlled, phase 3 clinical trials/studies are regarded as the most rigorous method of investigating drug efficacy and safety. However, due to their precise inclusion and exclusion criteria, these clinical studies frequently enroll a relatively small number of patients, which may not always reflect all possible consumers of the therapy. Clinical studies are also undertaken for a short period of time, making ADEs with a lengthy latency difficult to identify. Furthermore, following regulatory clearance, drug labeling and/or prescribing practices may change to incorporate new indications or patient populations, off-label usage, or concurrent use with other pharmaceuticals. Each of these new variables may lead to the development of ADEs that were not previously reported during clinical trials. Even over-the-counter pharmaceuticals, such as nonsteroidal anti-inflammatory drugs and phenylpropanolamine, have been linked to documented adverse drug reactions (ADRs) following regulatory approval, resulting in product withdrawal or labeling modifications [22, 23].

Data mining medication safety record databases, medical literature, and other digital resources could be useful in supplementing information concerning ADEs gathered during short-term clinical studies. Data mining for Pharmacovigilance may also create an "early warning system" that can discover drug safety risks faster than existing approaches. For these reasons, the FDA, the pharmaceutical industry, and drug safety experts are all interested in data mining these sources for ADEs.

III. Results

In the first part of the data analysis, we look at six filters that cause the distribution of average age by gender to change the most. Fig. 1 above shows the patients who most affect the distribution of the patients' ages. Patients 737, 1234 and 64 are among the patients who most affect the distribution of age.

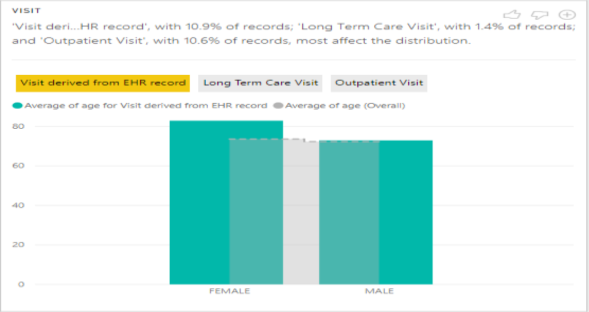


PERSON_ID

737, with 12.6% of records; 1234, with 13.8% of records; and 64, with 8.1% of records, among others, most affect the distribution.

Fig. 1. Affect the distribution of the patients' ages.

The visit derived from HR record, Long Term Care visit and Outpatient visit most affect the distribution of gender versus age at 10.9%, 1.4% and 10.6% of records respectively, as shown in Fig. 2.



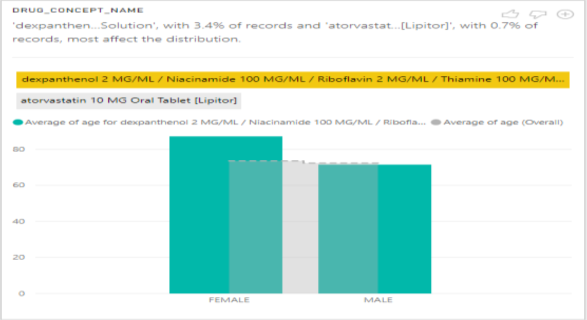
VISIT

'Visit deri...HR record', with 10.9% of records; 'Long Term Care Visit', with 1.4% of records; and 'Outpatient Visit', with 10.6% of records, most affect the distribution.

Visit derived from EHR record
Long Term Care Visit
Outpatient Visit
● Average of age for Visit derived from EHR record ● Average of age (Overall)
80
60
40
20
0
FEMALE
MALE

Fig. 2. The distribution of gender versus age.

Among the drug concept names, Dexpanthenol and Atorvastat are the drugs that most affect the distribution of the age and gender, as shown in Fig. 3.



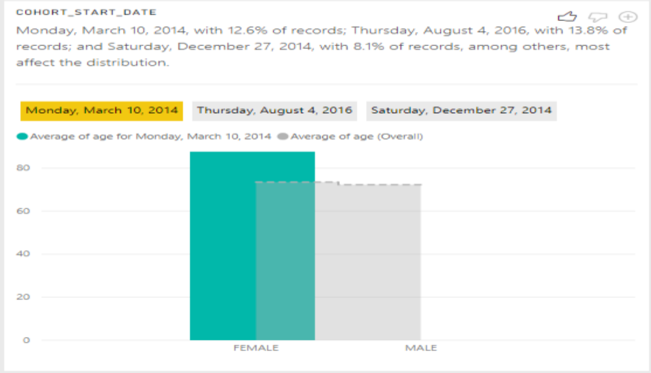
DRUG_CONCEPT_NAME

'dexpanthen...Solution', with 3.4% of records and 'atorvastat...[Lipitor]', with 0.7% of records, most affect the distribution.

dexpanthenol 2 MG/ML / Niacinamide 100 MG/ML / Riboflavin 2 MG/ML / Thiamine 100 MG/M...
atorvastatin 10 MG Oral Tablet [Lipitor]
● Average of age for dexpanthenol 2 MG/ML / Niacinamide 100 MG/ML / Ribofla... ● Average of age (Overall)
80
60
40
20
0
FEMALE
MALE

Fig. 3. The distribution of the age and gender.

Among the Cohort start dates, Monday, March 10, 2014, with 12.6% of records; Thursday, August 4, 2016 with 13.8% of records and Saturday, December 27, 2014 with 8.1% of records, among others, most affect the distribution of gender and age, as shown in Fig. 4.



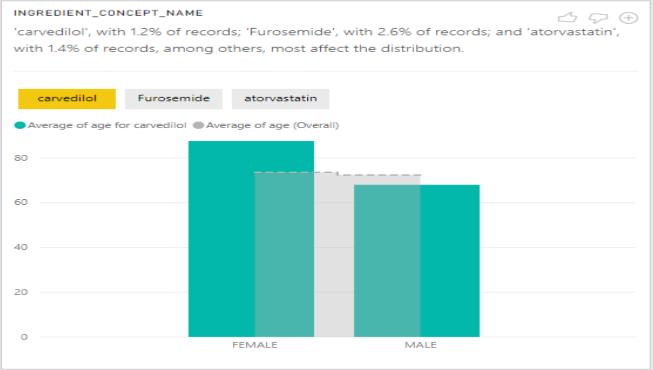
COHORT_START_DATE

Monday, March 10, 2014, with 12.6% of records; Thursday, August 4, 2016, with 13.8% of records; and Saturday, December 27, 2014, with 8.1% of records, among others, most affect the distribution.

Monday, March 10, 2014
Thursday, August 4, 2016
Saturday, December 27, 2014
● Average of age for Monday, March 10, 2014 ● Average of age (Overall)
80
60
40
20
0
FEMALE
MALE

Fig. 4. The distribution of gender and age on March 10, 2014, August 4, 2016 and December 27, 2014.

Among the ingredient concepts, Carvedilol, Furosemide and Atorvastatin most affect the distribution of gender and age comparison at 1.2%, 2.6% and 1.4% respectively, as shown in Fig. 5.

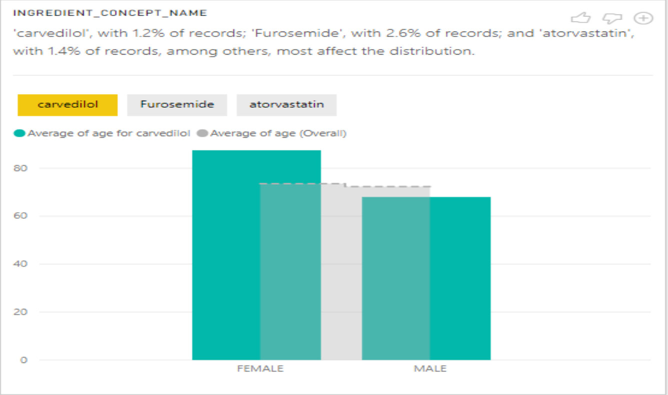


INGREDIENT_CONCEPT_NAME

'carvedilol', with 1.2% of records; 'Furosemide', with 2.6% of records; and 'atorvastatin', with 1.4% of records, among others, most affect the distribution.

Fig. 5. Effect of the distribution of gender and age.

Among the ingredient concepts, Carvedilol, Furosemide and Atorvastatin most affect the distribution of gender and age comparison at 1.2%, 2.6% and 1.4%, respectively, as shown in Fig. 6.

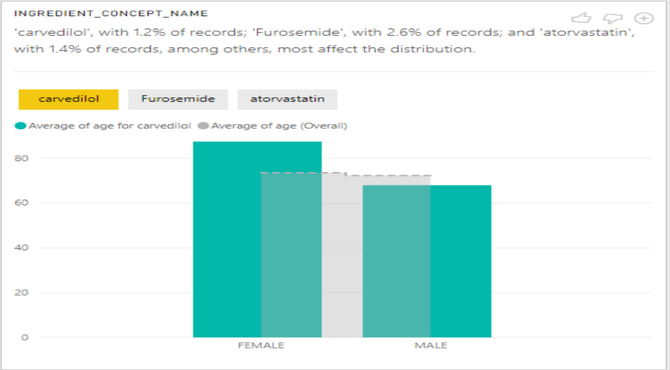


INGREDIENT_CONCEPT_NAME

'carvedilol', with 1.2% of records; 'Furosemide', with 2.6% of records; and 'atorvastatin', with 1.4% of records, among others, most affect the distribution.

Fig. 6. Distribution of gender versus age comparison.

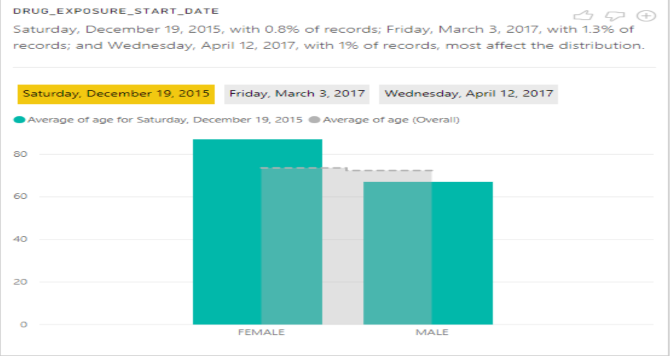
Among the ingredient concepts, Carvedilol, Furosemide and Atorvastatin most affect the distribution of gender and age comparison at 1.2%, 2.6% and 1.4% respectively, as shown in Fig. 7.


INGREDIENT_CONCEPT_NAME

'carvedilol', with 1.2% of records; 'Furosemide', with 2.6% of records; and 'atorvastatin', with 1.4% of records, among others, most affect the distribution.

Fig. 7. Ingredient concepts Carvedilol, Furosemide and Atorvastatin.

Among the drug exposure start dates, Saturday, December 19, 2015, with 0.8% of records; Friday, March 3, 2017, with 1.3% of records and Wednesday, April 12, 2017, with 1% of records most affect the distribution, as shown in Fig. 8.


DRUG_EXPOSURE_START_DATE

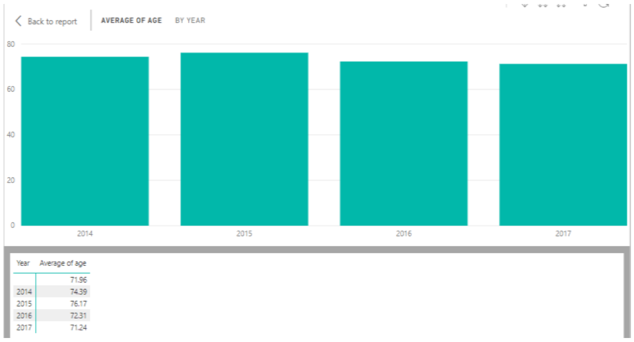
Saturday, December 19, 2015, with 0.8% of records; Friday, March 3, 2017, with 1.3% of records; and Wednesday, April 12, 2017, with 1% of records, most affect the distribution.

Saturday, December 19, 2015
Friday, March 3, 2017
Wednesday, April 12, 2017
● Average of age for Saturday, December 19, 2015 ● Average of age (Overall)
80
60
40
20
0
FEMALE
MALE

Fig. 8. The drug exposure starts dates.

A. Age of Patients Compared with Drug Exposure End Date

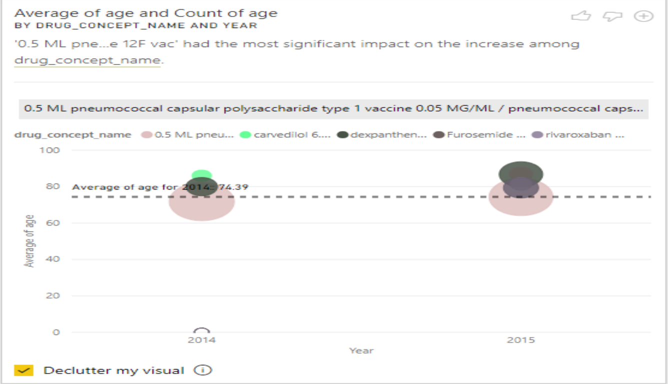
The year 2015 seems to have the oldest patients at the drug exposure end date at an average of 76 years old, as shown in Fig. 9.



Year	Average of age
	71.96
2014	74.39
2015	76.17
2016	72.31
2017	71.24

Fig. 9. Age of patients compared with drug exposure end date.

Fig. 10 shows the analysis of the 2.39% increase in average age between 2014 and 2015.


Average of age and Count of age

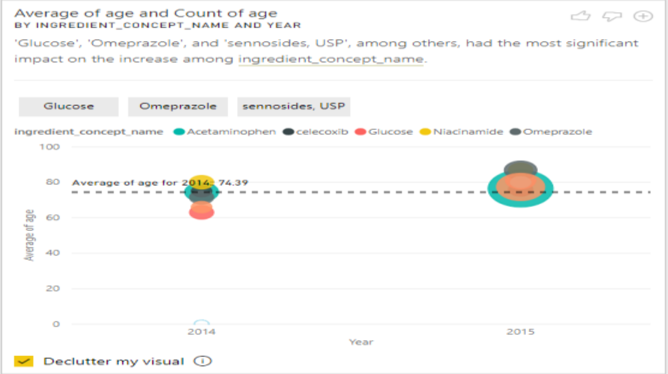
BY DRUG_CONCEPT_NAME AND YEAR

'0.5 ML pne...e 12F vac' had the most significant impact on the increase among drug_concept_name.

0.5 ML pneumococcal capsular polysaccharide type 1 vaccine 0.05 MG/ML / pneumococcal caps...
drug_concept_name ● 0.5 ML pneu... ● carvedilol 6.... ● dexpanthen... ● Furosemide ... ● rivaroxaban ...
100
80
60
40
20
0
Average of age for 2014: 74.39
Average of age
2014
2015
Year
Declutter my visual

Fig. 10. Average age between 2014 and 2015.

An analysis of drug concept versus age compared by year shows that 0.5 ML pneumococcal had the most significant impact on the increase among drug concept names, as shown in Fig. 11.


Average of age and Count of age

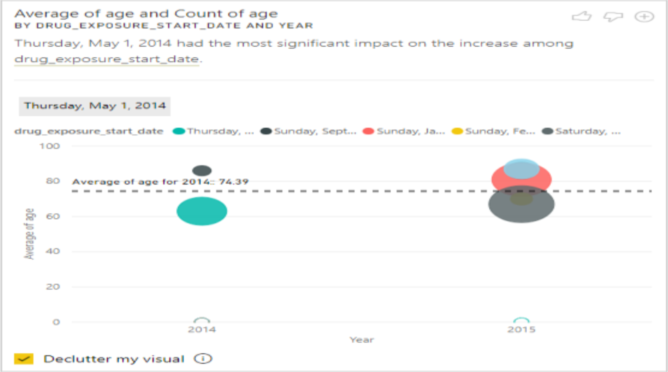
BY INGREDIENT_CONCEPT_NAME AND YEAR

'Glucose', 'Omeprazole', and 'sennosides, USP', among others, had the most significant impact on the increase among ingredient_concept_name.

Glucose
Omeprazole
sennosides, USP
ingredient_concept_name ● Acetaminophen ● celecoxib ● Glucose ● Niacinamide ● Omeprazole
100
80
60
40
20
0
Average of age for 2014: 74.39
Average of age
2014
2015
Year
Declutter my visual

Fig. 11. Analysis of drug concept versus age.

Glucose, Omeprazole and Sennosides had the most significant impact on the increase among ingredient concept names, as shown in Fig. 12.


Average of age and Count of age

BY DRUG_EXPOSURE_START_DATE AND YEAR

Thursday, May 1, 2014 had the most significant impact on the increase among drug_exposure_start_date.

Thursday, May 1, 2014
drug_exposure_start_date ● Thursday, ... ● Sunday, Sept... ● Sunday, Ja... ● Sunday, Fe... ● Saturday, ...
100
80
60
40
20
0
Average of age for 2014: 74.39
Average of age
2014
2015
Year
Declutter my visual

Fig. 12. Effect of the Glucose, Omeprazole and Sennosides.

Thursday, May 1, 2014, had the most significant impact on the increase among drug exposure start date. Table I above shows that visits derived from EHR record make up 10.9% of all patients followed closely by outpatient visits at 10.6%. Women are the majority at 52% of all the patients, as shown in Table II.

TABLE I. VISITS DERIVED FROM EHR RECORD

Type of visit	Frequency	Percent
Missing	**1499**	**75.9**
Emergency Room Visit	**15**	**0.8**
Inpatient visit	**8**	**0.4**
Long term care visit	**27**	**1.4**
Outpatient visit	**209**	**10.6**
Visit derived from HER record	**216**	**10.9**
Total	**1974**	**100**

TABLE II. FREQUENCY OF GENDER

Gender	Frequency	Percent
Female	**1017**	**51.5**
Male	**957**	**48.5**
Total	**1974**	**100**

The Cramer's V value measures the strength of the association between two categorical variables. The value of Cramer's V in this test is 0.97 which implies that there is a strong association between the cohort start date and the drug exposure start date, as shown in Table III.

TABLE III. ASSOCIATION BETWEEN THE COHORT START DATE AND THE DRUG EXPOSURE START DATE

Symmetric Measures			
		Value	***Approximate Significance***
Nominal by Nominal	Phi	6.0627918	0
	Cramer's V	0.9708237	0
N of valid cases		1974	

The Cramer's V value measures the strength of the association between two categorical variables. The value of Cramer's V in this test is 0.74 which implies that there is a strong association between the type of hospital visit and the type of drug administered, as shown in Table IV.

TABLE IV. ASSOCIATION BETWEEN THE TYPE OF HOSPITAL VISIT AND THE TYPE OF DRUG ADMINISTERED

Symmetric Measures			
		Value	***Approximate Significance***
Nominal by Nominal	Phi	1.6622077	0
	Cramer's V	0.7433619	0
N of valid cases		1974	

The Cramer's V value measures the strength of the association between two categorical variables. The value of Cramer's V in this test is 0.18 which implies that there is a weak association between the gender of a patient and the type of hospital visit, as shown in Table V.

TABLE V. ASSOCIATION BETWEEN THE GENDER OF A PATIENT AND THE TYPE OF HOSPITAL VISIT

Symmetric Measures			
		Value	***Approximate Significance***
Nominal by Nominal	Phi	0.1789966	0
	Cramer's V	0.1789966	0
N of valid cases		1974	

B. Discussion

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

Results from the data analysis show that patients 737, 1234 and 64 are among the patients who most affect the distribution of age. When we look at the purpose of the hospital visit, the visits derived from HR record, long term care visit and outpatient visits most affect the distribution of gender versus age at 10.9%, 1.4% and 10.6% of records respectively. Among the drug concept names, Dexpanthenol and Atorvastat are the drugs that most affect the distribution of the age and gender.

As for the Cohort start dates, Monday, March 10, 2014, with 12.6% of records; Thursday, August 4, 2016, with 13.8% of records and Saturday, December 27, 2014, with 8.1% of records, among others, most affect the distribution of gender and age. Among the ingredient concepts, Carvedilol, Furosemide and Atorvastatin most affect the distribution of gender and age comparison at 1.2%, 2.6% and 1.4%, respectively.

The year 2015 seems to have the oldest patients at the drug exposure end date at an average of 76 years old. There was a 2.39% increase in the average age of all patients between 2014 and 2015. An analysis of drug concept versus age compared by year shows that 0.5 ML pneumococcal had the most significant impact on the increase among drug concept names. Glucose, Omeprazole and Sennosides had the most significant impact on the increase among ingredient concept names.

Thursday, May 1, 2014, had the most significant impact on the increase among drug exposure start date. The visits derived from EHR record make up 10.9% of all patients followed closely by outpatient visits at 10.6%. Women are the majority at 52% of all the patients. Most females' visit was derived from EHR record while majority of the males visited the hospital for outpatient services.

There is a statistically significant relationship between the cohort start date and the drug exposure start date. The p-value is less than 0.05 and we therefore reject the null hypothesis and conclude that there is an association between the cohort start date and the drug exposure start date. The Cramer's V value measures the strength of the association between two categorical variables. The value of Cramer's V in this test is 0.97 which implies that there is a strong association between the cohort start date and the drug exposure start date.

There is a statistically significant relationship between the type of hospital visit and the type of drug administered. The p-value is less than 0.05 and we therefore reject the null hypothesis and conclude that there is an association between the type of hospital visit and the type of drug administered. The value of Cramer's V in this test is 0.74 which implies that there is a strong association between the type of hospital visit and the type of drug administered.

There is a statistically significant relationship between the gender of a patient and the type of hospital visit. The p-value is less than 0.05 and we therefore reject the null hypothesis and conclude that there is an association between the gender of a patient and the type of hospital visit. The value of Cramer's V in this test is 0.18 which implies that there is a weak association between the gender of a patient and the type of hospital visit.

IV. Conclusion

In conclusion, it is clear that there is a statistically significant association between several variables in our sample data. For instance, there is a fairly strong association between the type of hospital visit and the type of drug administered, which could imply that doctors prescribe medicine based on the type of visit to the hospital by the patient. However, it is worth noting that correlation or association between two variables does not imply causality and we recommend further research to delve deeper into the insights garnered by this study.

References

[1] Reps J et al. Investigating the Detection of Adverse Drug Events in a UK General Practice Electronic Health-Care Database. Arxiv; 2013: 1-7.

[2] Morimoto T et al. Adverse drug events and medication errors: detection and classification methods. QualSaf Health Care. 2004 ;13(4):306-14

[3] Häyrinen K et al. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform. 2008 ;77(5):291-304

[4] Wang X et al. Research Paper: Active Computerized Pharmacovigilance Using Natural Language Processing, Statistics, and Electronic Health Records: A Feasibility Study. 2009. JAMA; 16: 1-10.

[5] Aramaki E et al. Extraction of adverse drug effects from clinical records. Stud Health Technol Inform. 2010;160(Pt 1):739-43

[6] Kroeze JH, Matthee MC, Bothma TJD. Differentiating data- and text-mining terminology; Proceedings of the 2003 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology. 954024: South African Institute for Computer Scientists and Information Technologists; 2003. pp. 93–101.

[7] Witten IH. "text mining". In: Singh MP, editor. Practical handbook of internet computing. Boca Raton, Florida: Chapman & Hall/CRC Press; 2005. 14-1 - -22.

[8] Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf. 2002; 25(6):381–392. [PubMed]

[9] Harpaz R, Dumouchel W, Lependu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of Pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clinical pharmacology and therapeutics. 2013; 93(6):539–546. [PMC free article] [PubMed]

[10] DuMouchel W. Multivariate Bayesian Logistic Regression for Analysis of Clinical Study Safety Issues. Statist Sci. 2012; 27(3):319–339.

[11] Honig PK. Advancing the science of Pharmacovigilance. Clinical pharmacology and therapeutics. 2013; 93(6):474–475. [PubMed]

[12] 12. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clinical pharmacology and therapeutics. 2012; 91(6):1010–1021. [PMC free article] [PubMed]

[13] Prescription Drug User Fee Act (PDUFA V) [Accessed Apr 2014]; http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm272170.htm.

[14] REGULATION (EU) No 1235/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. [Accessed Apr 2014];2010 Dec

[15] Food and Drug Administration Amendments Act (FDAAA) of 2007. [Accessed Apr 2014];http://www.fda.gov/regulatoryinformation/legislation/federalfooddrugandcosmeticactfdcact/significantamendmentstothefdcact/foodanddrugadministrationamendmentsactof2007/default.htm.

[16] Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The New Sentinel Network - Improving the Evidence of Medical-Product Safety. New England Journal of Medicine. 2009; 361(7):645–647. [PubMed]

[17] Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. Journal of the American Medical Informatics Association. 2011;18(5):668–674.[PMC free article] [PubMed]

[18] Boyce RD, Ryan PB, Noren GN, et al. Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. [2014/07/02]; Drug Safety. 2014:1–11. [PMC free article] [PubMed]

[19] Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang C-K, Jackson RG, et al.Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC medical informatics and decision making. 2013; 13(1):71. [PMC free article] [PubMed]

[20] Vilar S, Friedman C, Hripcsak G, et al. Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature, and social media [published online February 17, 2017] Brief Bioinform. doi: 10.1093/bib/bbx010. [PubMed]

[21] Coloma PM, Trifiro G, Patadia V, Sturkenboom M. Post-marketing safety surveillance: Where does signal detection using electronic health care records fit into the big picture? Drug Saf. 2013; 36:183–197. [PubMed]

[22] Food and Drug Administration. FDA issues public health warning on phenylpropanolamine.[Accessed September 22, 2017].

[23] Cantu C, Arauz A, Murillo-Bonilla LM, et al. Stroke associated with sympathomimetics contained in over-the-counter cough and cold drugs. Stroke. 2003; 34(7):1667–1672. [PubMed]