Adverse Drug events and Pharmaceutical Company profits*

Adam Tammariello
University of Colorado Boulder
Boulder, Colorado

Taicheng Song
University of Colorado Boulder
Boulder, Colorado

ACM Reference Format:
Adam Tammariello and Taicheng Song. 2018. Adverse Drug events and Pharmaceutical Company profits. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, Article 4, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 MOTIVATION

The Federal Food and Drug Association provides data on the outcomes of specific adverse events pertaining to patients who report such incidents. Since the health care industry in the United States is a widely debated topic, it would be interesting to find specific data about Pharmaceutical companies providing drugs that have severe outcomes. With the combination of stock reports and profit margins, compared to the data provided by the FDA, it would be interesting to draw specific correlations between profit margins and years that the FDA reported high mortality rates related to a drug, or other levels of severity. The hope would be to provide specific reports that would contribute to the discourse of privatized health care.

2 LITERATURE SURVEY

Prior Work done within this topic have shown to correspond to prescription drugs and drug abuse in and out side the United States. Theses studies are done with FDA Data and without.
In a Study done by Brooke Belong of the Odyssey, Health Care corporations circulate highly addictive drugs as prescription drugs, and draws a correlation between these prescriptions and increasing drug abuse rates within the country. These statistics are drawn from the FDA.
https://www.theodysseyonline.com/legal-drugs-are-more-dangerous-than-illegal-ones
In a Study done by Donald W. Light, he draws a correlation of adverse outcomes from FDA reported new prescription drugs to pharmaceutical companies creating high risk prescription drugs to push large profit margins. This data incorporates FDA adverse Event reporting.
Harvard Study

*Produces the permission block, and copyright information

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

3 PROPOSED WORK

3.1 Data Collection

To begin this investigation, we will need to collect specific data from the FDA Adverse Event Reporting portal. This can be accessed either directly through the FDA website, or through the data collection website Engima.org. From here we can pick specific years, in this case we will choose a period of 3-5 years. This will depend on what access we have to specific stock and profit information of specific Pharmaceutical companies that we have access to. With a time frame decided we can then find the corresponding data for Drug Information within Adverse Event Reporting, as well as Patient outcomes with the corresponding Id's. Then once we have collected these two data sets, we find the stock information or profits of the specific corporation that has the highest rates of severity amongst there products or specific product.

3.2 Data Cleaning and Evaluation

To properly use the specific dataset, we will need to make sure we can process it easily to better provide visualizations. Within the FDA Adverse Event Reporting - Drug Information, and FDA Adverse Event Reporting - Patient Outcome, There are varying aspects to the data sets that can be cleaned. Since we are centralizing on levels of severity, data inputs that do not have any data corresponding to an outcome in Patient Outcome, will not be useful to us. Other non-specific descriptors of drugs as well as variations of missing descriptors. This allows us to easily

3.3 Data Preprocessing

To process the data after pruning unimportant data, we must change specific aspects of the data to be able to properly visualize it. Within the Patient Outcome dataset, the attribute "Outcome" comes with a small initial, as well as a description. For our use, we will assign each specific outcome to a specific number so as to understand the fequency of outcomes of specific drugs. This will make it easier to compare frequencies. We then combine the Patient Outcome Data set with it's corresponding ID number within the Drug information data set.

3.4 Data Evaluation

Finding the most frequently reported drug by using priori algorithm From the most frequently reported drugs, find the most frequent outcome (death, disability, minor adverse effect etc.) and looking at the most severe outcomes. Showing stock behavior of manufacturing company with the most severe output

3.5 Differences

In comparison to the prior work done, our work will be more generalized on how the public domain views pharmaceutical companies

during periods of high mortality or injury, as well as the profits they gain from high risk drugs. This will connect or our main goal, which is contributing to the discourse of privatized health care, as well corroborate previous studies done in the same field. Our study will only be for The United states as well, where as the Harvard study was more global. Since the United States is one of few countries with privatized health care, we can see how domestically based pharmaceuticals gain profit during the death of domestic drug related deaths.

4 DATA SET

4.1 U.S FDA - Adverse Event Reporting - Drug Information

This Data set contains over 1,020,344 data points pertaining to specific attributes. The overall purpose of the data set is to show specific case ID's of a specific adverse event, and the corresponding drug information used in that event. There are 14 Attributes in the set which have the following meaning:

- IDr - specific case ID
- Drug Sequence Identifier No.- Another identifier for specificity
- Drug Role - Whether or not the Drug is suspected to be the cause of the even
- Drug Name -Name of medicinal product
- Validated/Verbatim - Whether or not a trade name is used or a Verbatim name
- Route - method in which drug was taken
- Dose - how much of a drug was taken
- Other identifiers ·

The link to the Data can be found here: Drug Information

4.2 US FDA - Adverse Event Reporting - Patient Outcome

this data set contains specific outcomes for a corresponding identifiers from other tables within the Adverse Event reporting data collection. From here we can less data points, however there are multiple points within the Drug information Data set which have the same identifiers. These correspond with those specific drugs and their outcomes. This Data set has 4 attributes:

- Idr - Specific Identification number
- outcome - and abbreviation for an Outcome (Death, Disability, Hospitalization etc.)
- Outcome Definition - corresponding defintions to outcome abbreviations
- Quarter - which quarter of the year the event happened.

The link to the data can be viewed here: Patient Outcome

4.3 Historical Quotes - NASDAQ

This data set will be specified once the initial work has been put in to decipher specifically which drug has the most adverse affect, as well as the main distributor of said drug. From there we can pick the stock information and hopefully profits of that specific company during the time frame specified and look at the information provided to see a correlation. this data set has 4 attributes:

- Time - the time in which the quote was recorded
- Open - The price of stock at Open
- High - The highest sold stock
- Low - The lowest Sold Stock
- Close- Price at closing time
- Volume- the amount of transactions that happened at this time

The data can be viewed here: Pfizer Historical Quotes

5 EVALUTAION METHODS

Correlating our data will be fairly easy, the time consuming part will be processing our data to determine which specific drug has the most frequent severity reports. To do this we will use an apriori algorithm to determine the types of drugs that have the most sever outcomes. Then we will use this to determine which specific drug has either the wides ranging severity, or the most sever outcome by probability. From this list we will widdle down a small selection of drugs and their corresponding companies and find stock reports for each company during a specific time frame. We can find a correlation to each specific companies profits or stocks during a year with a high severity drug reporting. This will show us whether or not companies gain or lose profit if a drug has been shown to hospitalize patients or even kill them.

6 TOOLS

6.1 Excel

Using Excel we can visualize cvs files from the ouputs created by our data processing algorithms. This will help us check our work and even create simple visualizations of data that we have so far. this will be especially helpful in collaboration so that we can share results with eachother without having to run written programs in Jupyter and keep us organized when we finalize our results.

6.2 Python

In class we have had a lot of practice with the python programming language. This has proven to be especially efficient and easy to use to process data in very versatile ways. Using libraries such as numpy, pandas, cvs etc.. we can read in our data sets and clean our data sets of arbitrary information and even produce detailed visualizations of our frequency vs drug tables, as well as the correlations between profit rates and severe drugs. This will be our workbench for properly parsing data and sorting it.

6.3 Jupyter

Jupyter Lab and Jupyter Notebook will be very essential for the group to use python. This will create a basis in which regardless of the system, each member will be able to use and run python algorithms that we create. We can also use this to easily save files and push to our version control platform.

6.4 Github

We will be using github as our main source of version control. This will allow us to develop in tandem and contribute to a main product which will be the collection of data and visualizations as well as the programs we create to parse specific information. It also gives us

the ability to look at participation in the group as well as a timeline of our progress.

6.5 Slack/Trello

We will strive to use Slack as a means of communication to properly discuss topics and share specific files and resources. It will also allow us to know when a member of the group pushes anything to the github so that we can approve any pull requests. We may also use trello as a resource to keep track of what work is still needed and what work has been completed to keep on track with our milestones.

7 MILESTONE

- Data cleaning done by March 23rd
- Data Preprocessing done by April 10th
- FDA Dataset Evaluated by April 17th
- NASDAQ Pharma Data correlation by April 20th
- Presentation ready by April 27th

Work has already begin on data cleaning. However with spring break as well as different classwork we will try to prioritize the aspects of the work that will take longer than others. Most do these dates will be subject to change to adjust for different roadblocks along the way.

8 REVIEW

During our presentation we were given some critiques of our work thus far. Our research on the prior work done on our subject had been lacking since we only produced one specific article loosely pertaining to our subject. We have since then found more sources to draw inspiration and guidance so that we can use the Dataset and topic in a new and interesting way. We ran out of time during our presentation so our questions were limited, but for the critique we got this was what we improved on first.

REFERENCES