

# Observations and Analysis of Open Payments Database

Harper Schwab

Data 11800, Professor William Trimble

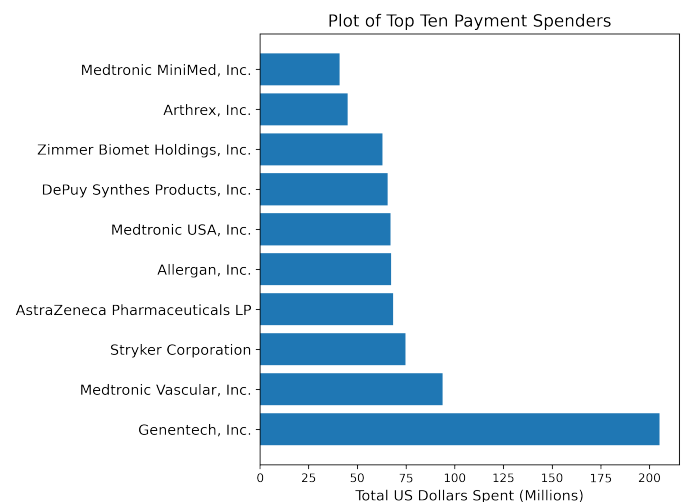
Due February 22, 2022

## Introduction and Dataset Background

The Open Payments database is a department of Health and Human Services operated and mandated database which is used to compile a report to Congress every year and available to the public. This paper uses data from both the 2019 [4] and 2020 [5] Open Payments Datasets. These findings are from the General payments subsets of the database, which contains payments that do not fall under the categories of research, or ownership. [3] This means that all of the payments analyzed are "Payments or other transfers of value made that are not in connection with a research agreement or research protocol". [3] The majority of these take the form of In-kind items and services, which are more commonly payments for Food and Beverage, compensation for speaking at an event, or travel. These payments are to doctors from pharmaceutical companies, with the goal of using the paid doctor's credibility to present information about their drugs to other doctors. The companies do this in order to increase sales of drugs that they have produced. The New York Times article [1] *Dr. Drug Rep*, provides first hand context describing a drug company making multiple payments in order to have a doctor, the author, speak at events and have private meetings with other doctors where he often stretched the truth about these drugs, which caused doctors, including himself, to prescribe a drug when they most likely shouldn't have unknowingly assisting the drug company in having higher sales. This proposes many ethical issues, hence the requirement of a yearly report to congress. However, it is still important to understand where the money comes from, who is receiving the money or, and how do the payments change over time. All of these questions help to demonstrate the importance and constant nature of many drugs advertisements and representation.

## Where does the money come from?

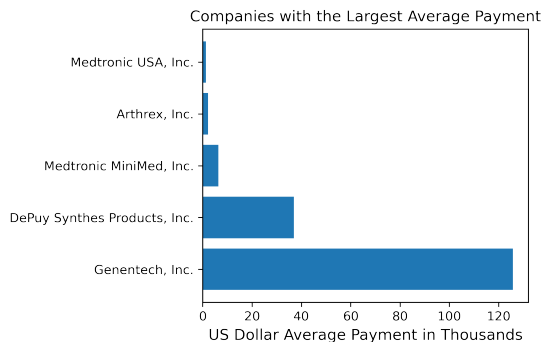
In order to identify where the money is coming from I created a subset of the whole 2020 dataframe [5], containing attributes important to identifying where the money is coming from. Since these payments are coming from pharmaceutical companies I wanted to identify the pharmaceutical companies with the highest payment amount and highest average payment amount. For the first plot created, the relevant attributes are the pharmaceutical company name, the total amount spent, and the number of payments. The values were then grouped by the pharmaceutical company, resulting in a dataframe with pharmaceutical company name and total US Dollar amount of payment. These values were then sorted in order of highest to lowest expenditure amount, sliced in order to only show the top 10, then plotted on a horizontal bar chart.



The figure above illustrates that the pharmaceutical company with the largest expenditure in 2020 was Genentech Inc. Which actually had \$111,327,310 more than the total payment amount than Medtronic Vascular Inc, the second highest spender of 2020.

I then thought it would be interesting to examine companies to with the largest average payment to

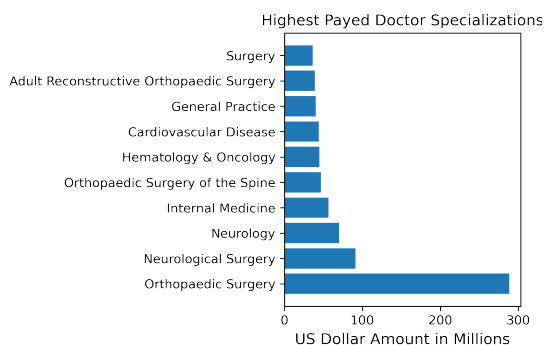
see if there is any overlap. To do this I used the grouped values from the previous figure, and then divided the total dollar amount of payments by the total number of payments.



Above we can see that Genentech INC is also the highest average spending company per payment. This is most likely because they had a few extremely large payments, 102 payments above \$100,000 to be exact. These payments skewed their numbers compared to the other pharmaceutical companies which explains their significant increase compared to the other high payment companies in both the total and average payment amounts.

### Where does the money go?

To identify where the money specifically goes, I decided to investigate the specialization of the physicians in the dataset. In order to see what type of doctors are receiving the most money, I took a subset of the full 2020 dataset using the columns physician specialization and full payment amount. After grouping by the specialization and sorting them in descending order I shortened the labels and plotted a barh graph seen in the following plot.



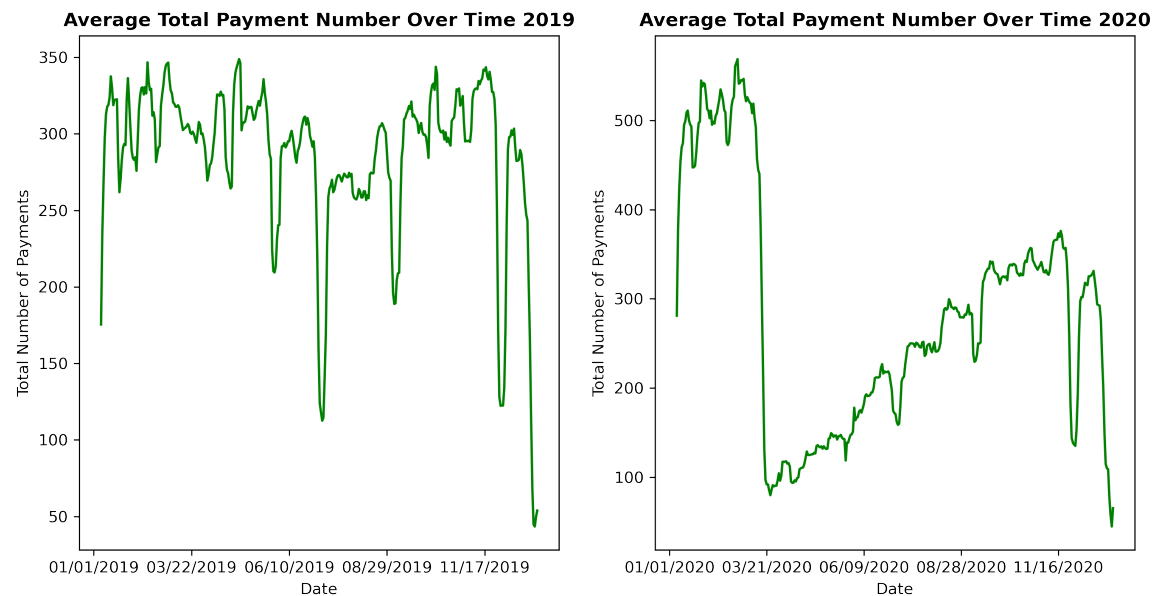
The plot above shows that the most common type of doctor that is paid by the pharmaceutical companies is an orthopaedic surgeon, followed by

neurological surgeon. Which demonstrates that the pharmaceutical companies are paying orthopaedic and neurological surgeons the most money in order to recommend and present their drugs to other doctors, as described in the New York Times Article [1]. This seems to demonstrate that many of the drugs that are peddled by the pharmaceutical companies are in orthopaedic or neurological fields. These are most likely painkillers, anti-inflammatory, and other drugs which help relieve negative orthopaedic symptoms because an orthopaedic doctor would one: know more about those specific drugs, and two: be more inclined to prescribe such drugs.

### How do things change?

When identifying a way these payments change, I decided to investigate the number of payments for two years: 2019, and 2020. For these plots, I took a random sample because using both of the full datasets caused an unreasonable run time on my computer. The samples from the 2019 [4] and 2020 [5] datasets contain 100,000 separate payments, and were created using the `df.sample` function from the pandas library. In order to format the data to create the plots I changed the Date of Payment column to datetime format, grouped the number of payments by the dates, sorted them from the beginning to the end of the year, then because it is the trend of the payment amount that will really convey the information I am looking for I calculated the 7 value rolling average of the payment numbers for the date values, and finally plotted them. I used the same method for both the 2019 and 2020 dataset.

As seen in the plot at the top of the next page, the COVID19 Pandemic greatly disturbed the normal trends of payment numbers. The occasional dips in the graph could be explained by the samples that were used for these calculations. I hypothesise that if I were to calculate this with the entire dataset there would not be seemingly random weeks in which there was a drop in payments greater than roughly 150 except for the expected decrease from the COVID19 Pandemic in 2020. On both graphs there is a decrease at the end of the year and a increase at the beginning of the new year. This could possibly be explained by doctors not presenting during the end-of-year holiday season, however this is just speculation.



## Conclusions

The Open Payments database provides an opportunity to shed light on the stories of these pharmaceutical companies paying physicians to advocate their product. Genentech Inc. a large biotech company based in San-Fransisco [2] had the largest total and average payments, these large payments may provide reason to investigate further into the nature of the payments and safety of their products. If Daniel Carlat's story [1] is analogous to the broad experience of these physicians turned drug representatives than payments like these should be investigated, because stretched statistics in medicine cause unnecessary symptoms and pain for patients [1]. With fields as important as orthopedics being the most common payed physicians, these findings also call into question the presence of biases of many orthopedic doctor's drug prescriptions. Will a doctor prescribe medicine based on stretched truths from a paid presentation with doctored slides like in the New York Times article [1]? On another topic, the frequency of these payments, shown on the previous figure, does not seem to be trending downwards. In the beginning of 2020 the average number of payments was higher than any point in 2019, there was the subsequent crash resulted from the COVID19 pandemic however it seems like the general trend will remain upwards, as towards the end of 2020 numbers reached the height of 2019's numbers. Meaning as a country we have not realized the danger of these payments by pharmaceutical companies.

## REFERENCES

- [1] Daniel Carlat. Dr. drug rep. *New York Times*, November.
- [2] Genentech. Genentech. date accessed: 02/21/22, /url<https://www.gene.com>.
- [3] Department of Health Human Services. Open payments program report to congress. April 2021. <https://www.cms.gov/files/document/open-payments-2020-annual-report-congress.pdf>.
- [4] [openpaymentsdata.cms.gov](https://openpaymentsdata.cms.gov). 2019 general payment data, 2019. data retrieved from open payments data, <https://openpaymentsdata.cms.gov/dataset/4e54dd6c-30f8-4f86-86a7-3c109a89528e>, updated: Updated January 21, 2022.
- [5] [openpaymentsdata.cms.gov](https://openpaymentsdata.cms.gov). 2020 general payment data, 2020. data retrieved from open payments data, <https://openpaymentsdata.cms.gov/dataset/a08c4b30-5cf3-4948-ad40-36f404619019>, updated: Updated January 21, 2022.