# Biotechnology Stock Analysis

## Python Project

Jason Lewris jl4pt@virginia.edu

Kaustav Shah ks3hw@virginia.edu

Yuanbo Wang yw4rd@virginia.edu

Sharath Pingula vp4pa@virginia.edu

**Introduction:**

The purpose of this research engagement was to use python to format, manipulate, and process data to provide insights regarding a biotechnology firm's stock performance. We wanted to see if there was any correlation between clinical trial results for a drug being sponsored by a biotechnology firm, and that firm's individual stock performance. This was to be looked at across large biotechnology firms, market capitalizations of approximately $100B, and small biotechnology firms, market capitalizations of approximately $1B.

In order for a drug to be approved for general consumer purchasing, it must undergo several phases of clinical trials. The announcement of results for these clinical trials were believed to have significant impacts on stock performance for the corresponding biotechnology firm.

In order to simplify the scope of the project, a list of 24 clinical trial results that were keyed as 'positive' or 'negative' from a research project in 2013 were used to test the corresponding companies stock performance for large biotechnology firms. For small biotechnology firms, clinical trial results posted within the past 2 months were used for analysis.

**The Data:**

The report used to determine whether a clinical trial result was positive or negative was completed in 2013 by a researcher named Thomas J. Hwang[1]. This research is hosted by journals.plos.org. The data was presented in tabular form, and was available for download. This spreadsheet included 24 observations and 8 variables. These variables included the event number, the company name, the expected sign, the phase, the indication, the compound, and the trial ID. The companies that were selected for this project include: Amgen, Gilead Sciences Inc., Pfizer, Merck, Biogen, Bristol-Myers Squibb Co, and Eli Lilly and Co. The dates of the project span Jan 1, 2011 to May 31, 2013.

In order to see the corresponding stock information for these companies during the span of these result dates, historical stock price information had to be downloaded. For each of the companies mentioned above, over the range of Jan 1, 2011 to May 31, 2013 daily stock price information was extracted from Google Finance. This data came preformatted, cleaned, and ready for ingestion into the python program. The data was stored in a CSV, and contained 4243 observations and 7 variables which included stock symbol, date, open price, high price, low price, close price, and volume of shares being traded.

**The Process:**

The historical stock price information was read into Python using Pandas. This data was stored as a data frame. The key dates referenced in the clinical trials report were used as the baseline date for the corresponding stocks. The event date was day 0, and stock price information for that particular company was extracted 8 days before and 8 days after the key event date.

This was done by sub setting the original historical stock price dataset by the company of interest, and then converting the date variable into a Python datetime variable. Pandas includes a nice feature called DateOffSet which enables a user to extract dates based on a number. So utilizing the DateOffSet features, -8 days from the key date and +8 days from the key date were extracted from the original historical stock price csv. This data was then stored in a separate data frame and written to a csv file.

Essentially there were 24 queries that took place, one for each key event date for the particular company that was participating in that clinical trial. For example, event 1 was on March 19, 2013 for Amgen. All 24 of these queries were done utilizing a method in python, which included the event date, the event number, and the stock symbol. From there a simple for loop was utilized to feed this method all 24 key event dates. At the end, a new csv file was created with stock price information covering a range of dates from the key event date, both before and after.

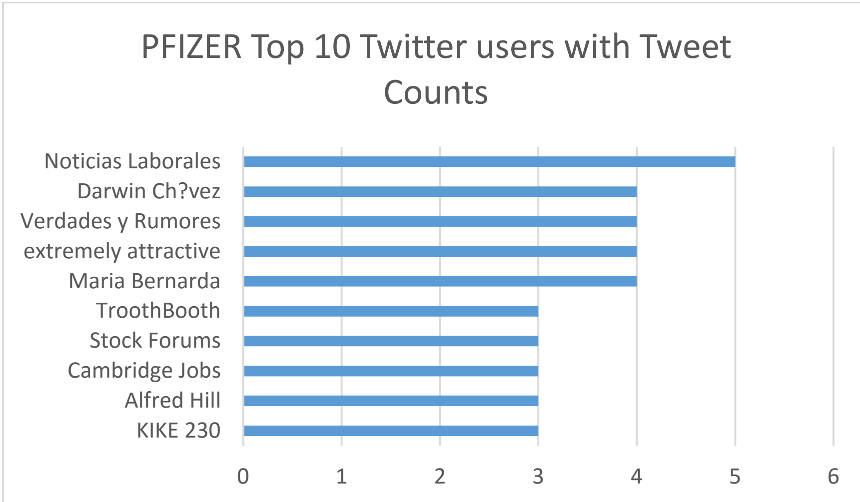
**Additional Features:**

Ideally, the python program created would be able to function with live data and handle more inputs. One input that would be of great value would be social media, such as Twitter and the number of times a particular biotech firm was mentioned. Further analysis could be done on this Twitter information, such as what words were being mentioned most frequently, and whether or not they were having an impact on the companies stock performance.

Unfortunately, Twitter makes it very difficult with their current API to extract historical information. Instead, to make this feature a reality, only the past 8 days worth of Twitter data was extracted, which was the limit of Twitters API. These tweets were extracted solely utilizing Python. This module that was created to connect to Twitters API and extract tweets were outside the original specifications of the project.

Tweets were extracted for all companies of interest over the past 8 days. Text analysis was then conducted on these tweets using python to count the frequency in which words occurred. This same analysis was conducted on the actual twitter accounts that were posting. The final result was the top ten words that occurred for each biotechnology firm, as well as the top ten contributors for each firms. An example output of the results can be seen below in figure 1.

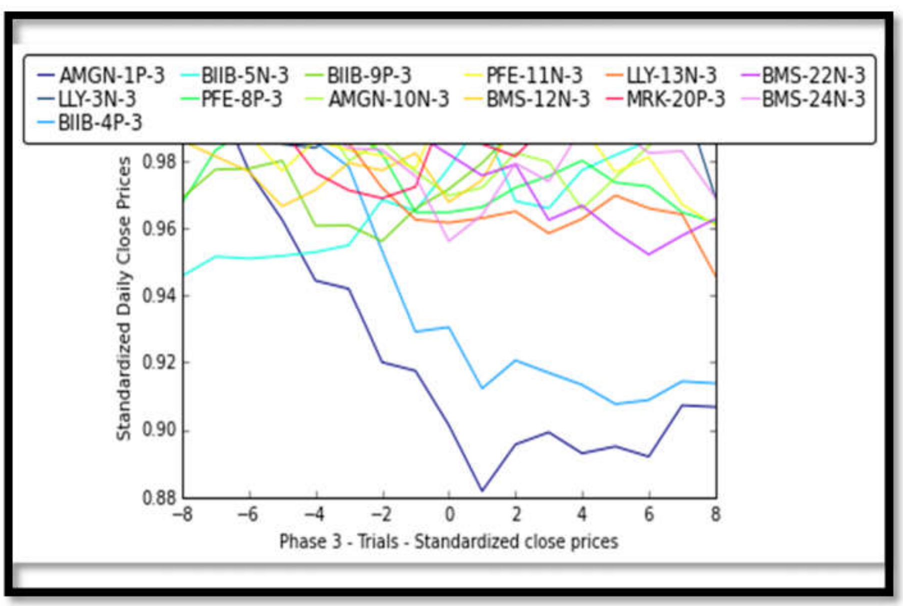
Figure 1




**Results:**

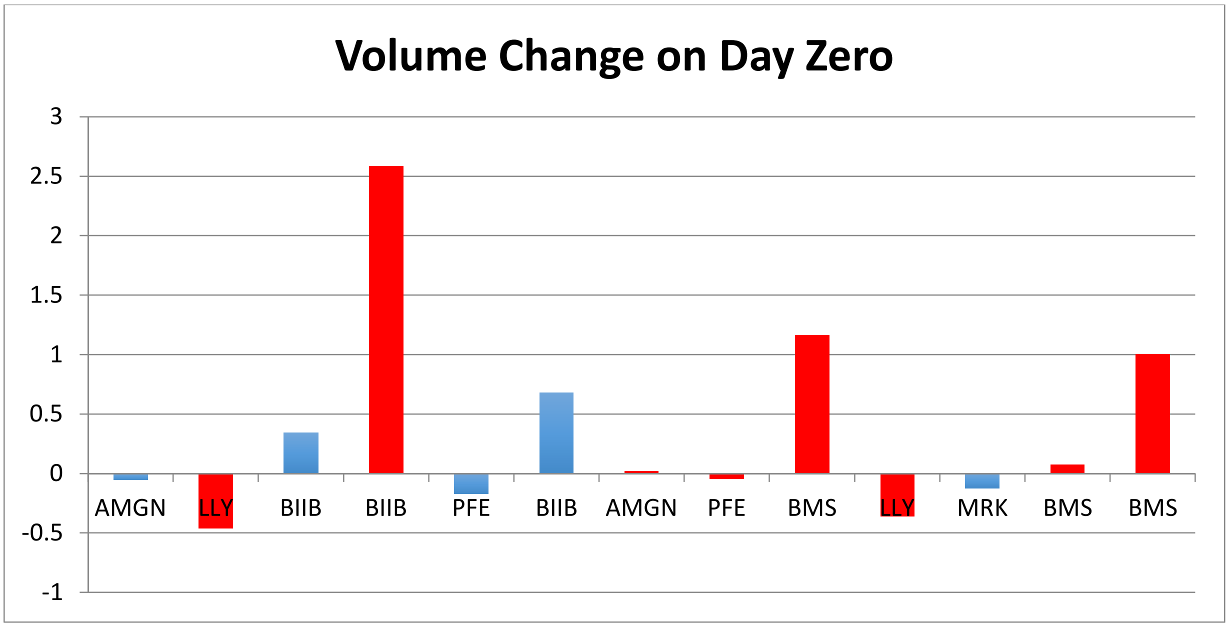
As can be seen in figure 1.1 below, the results were across the board regarding the impact these clinical trial postings had on stock performance for large biotechnology firms. Most stock prices hovered within a 5% range. There were however, two particular events that corresponded to fluctuations within stock performance of roughly 12%.

Figure 1.1



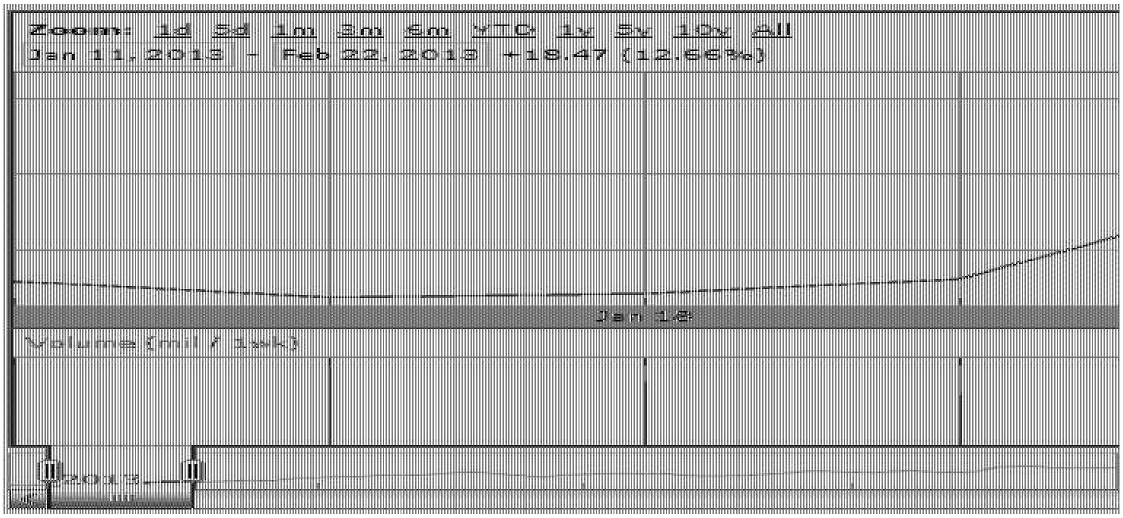
Taking a look at figure 1.2 below, we can see that there are significant changes in the volume of shares being traded for these large market cap biotechnology firms. Large changes in volume are often an indicator that institutional investors are either moving into or out of a particular company. One such company that had significant changes in the volume of shares being traded was Biogen (BIIB).

Figure 1.2



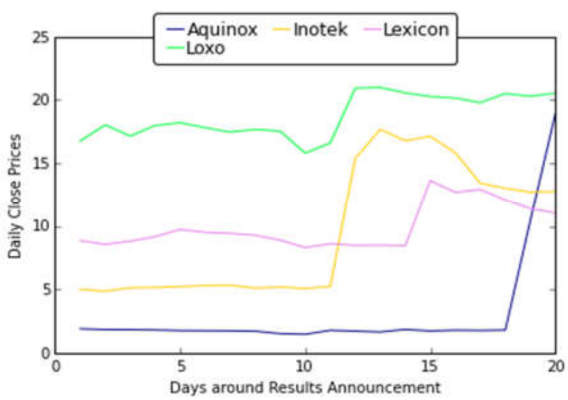

Since these are larger firms, the volume changes often take longer to be reflected in the actual stock price of the company. Continuing with the example of Biogen, figure 1.3 demonstrates the fluctuation in stock performance taking place within 30 days of the key event date transpiring, which was January 24, 2013.  This was outside the range of dates originally extracted in our initial analysis.

Figure 1.3

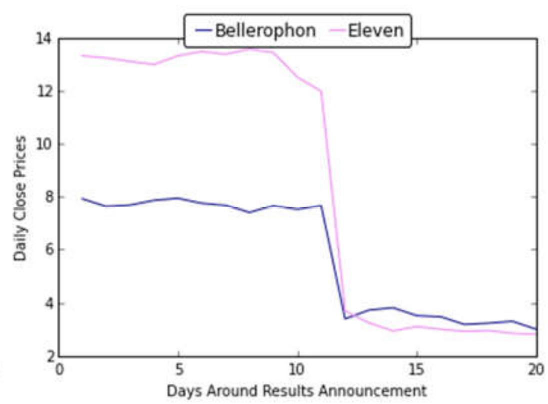



Lastly we wanted to take a look at the impact clinical trial results had on small market capitalization biotechnology firms. These are firms that have a market cap of approximately $1B. Looking at the figures below, we can see that the impact takes place almost immediately, and that these fluctuations in stock performance can be quite significant. In virtually all cases, there were significant changes in the stock price of a company within a very short period of time, often less than 2 or 3 days.

**Positive Results**




**Negative Results**




**Conclusion:**

There are several key takeaways from the results outlined above. The first of which is clinical trial results, both positive and negative, impact large market cap and small market cap biotechnology firms, but in different ways. For larger biotech firms, this impact immediately takes place in the volume of shares being traded. Since larger companies are often less sensitive to individual results, this translates into stock performance fluctuations within a longer period of time, roughly 30 days.

For small market cap biotechnology firms, clinical trial results impact stock performance rapidly, and these fluctuations can be quite significant. These fluctuations occur within several days of a result posting.

This type of information can be used by investors of all types to monitor and potentially invest in large and small biotechnology firms. Knowing that large firms often take longer for stock performance to be impacted, an investor could anticipate holding onto a particular stock for several months. For smaller biotechnology firms, investors can potentially see results within several days of a clinical trial announcement.

If more time was allowed, this project could have been improved in several ways. The first of which is instead of using historical data, the data streams coming in would be transitioned to live data. This includes stock prices, clinical trial postings, and social media reports. Additionally, the social media analysis could be more heavily integrated within the overall project.

Sources:

1. Hwang, T. J. (n.d.). Stock Market Returns and CLincial Trial Results of Investigational Compounds: An Event STudy Analysis of Large Biopharmaceutical Companies. Retrieved from Plos One website: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0071966