Abstract

I hypothesized that the social media platform Twitter is a good predictor of the monthly cancer incidence and mortality numbers issued by the National Cancer Institute. My project counted the number of tweets that contained words related to incidence or mortality and mentioned one of the seven (7) most common cancers. Therefore, for each month, twitter “incidence” and “mortality” data was collected for lung, breast, prostate, melanoma, colon, bladder and leukemia cancers and compared to monthly 2019 NIH incidence and mortality rates. To perform this analysis I generated an autonomous computer system that collected real-time cancer-related tweets each month. Using data analysis I then compared twitter “incidence” and “mortality” data to NIH data to see if they show similar trends or unique differences. This analysis has been running for the last 7 months and will be complete after one year has been collected. My analysis has revealed that in general twitter data can predict monthly cancer incidence and mortality rates, however, several months contained twitter data for specific cancers that was significantly different than the NIH data. These differences, in part, may provide insight into the effectiveness of cancer awareness campaigns or could uncover unique events that have occurred in that month as it relates to a specific cancer. This study will reflect on both of these scenarios and continues to show how Twitter is a powerful predictor of cancer statistics and events.

Methods

Ifttt searches the information off twitter puts on google sheets to be compared against NIH mortality and incident statistics

Incident Rate

Mortality Rate

Conclusion

- Tweets from twitter about cancer incidence and mortality generally mirror NIH cancer statistics
- Certain months for specific cancers show spikes in the amount of tweets
- Tweets about breast cancer spiked in October (breast cancer awareness month)
- Tweets about Leukemia spiked from November through March (reason unknown)
- Twitter can be used to discover and/or generalize events centered around specific cancers

I will be furthering this project for a complete year (until september 2021). Plus, I will compare it with NIH 2020 instead of NIH 2019.

I acknowledge my mentor Dr.Fata for the help of analyzing the data.

Acknowledgements