Public Sentiment about Medicare, Medicaid, and State-Based Medicaid Systems on Twitter

Avery Lumeng
POL1 317
20 December 2019
Thomas Burke

Many health care researchers argue that Medicaid, a program primarily for poor people, is stigmatized relative to Medicare, a program that covers seniors. Researchers from the University of Chicago assert, "Medicaid has long been considered Medicare's poor stepchild with little political support". [1]This means that, while Medicare is an oft-discussed, well-known policy with a clear political constituency, Medicaid is often less politically salient. MIT's Andrea Campbell also states that social assistance programs for the poor, including Medicaid, are "heavily stigmatized and so most of the eligible don't sign up". [2] In other words, Campbell believes that, even outside of the political-policy sphere, Medicaid is socially stigmatized leading to low enrollment numbers.

However, this stigmatization of Medicaid is not necessarily represented in public polling data. It is true that many surveys have found strong historical support for Medicare and specifically the benefits it provides; seven in ten Americans have "a favorable view of the program" as a whole according to one survey.[3] Regarding Medicaid, however, the conventional wisdom may not hold up. While research on public opinion of Medicaid is less robust, one poll from the Kaiser Family Foundation found that 74 percent of those polled have a "very" or "somewhat favorable" view of Medicaid;[4] further, 61 percent believe that Medicaid is "working well" in the nation, 67 percent in their own state,[5] and many surveys

---

[1] Colleen M Grogan and Sunggeun Ethan Park, "The Politics of Medicaid: Most Americans Are Connected to the Program, Support Its Expansion, and Do Not View It as Stigmatizing," *The Milbank quarterly* vol. 95,4 (2017): 749-782. doi:10.1111/1468-0009.12298

[2] Andrea Campbell, *Trapped in the Safety Net: One Family's Journey through America's Social Policy Programs*, 2014, Chicago: University of Chicago Press, Ch. 3, p 3.

[3] E.C. Hamel, C. Deane, and M. Brodie, 2011, "Medicare and Medicaid," in *American Public Opinion and Health Care*, Washington D.C.: CQ Press.

[4] Kaiser Family Foundation, 2017, "In general, do you have a favorable or unfavorable opinion of Medicaid?" June 14-19, 2017, *Kaiser Family Foundation Health Tracking Poll*, retrieved from https://www.kff.org/medicaid/poll-finding/data-note-10-charts-about-public-opinion-on-medicaid/.

[5] Kaiser Family Foundation, 2017, "Would you say the current Medicaid program is working…?" June 14-19, 2017, *Kaiser Family Foundation Health Tracking Poll*, retrieved from https://www.kff.org/medicaid/poll-finding/data-note-10-charts-about-public-opinion-on-medicaid/.

prior to passage of the Affordable Care Act (ACA) show strong support for expanding Medicaid as a means of reforming the health care system.[6]

Medicaid is a program partially run by the federal government and partially run by the states. Often, states name their Medicaid programs differently from the national program, potentially obfuscating the connection between the state-based program and the national Medicaid program. For example, the average Californian may not know that "Medi-Cal" is simply California's version of Medicaid, the average Tennessean that "TennCare" is just Medicaid in Tennessee. Research suggests that Medicaid might be even more popular when it is framed as a state rather than federal program. A number of polls have suggested that a growing majority of Americans prefer their state rather than their federal government, feeling as if they "get the most for [their] money"[7] and prefer to concentrate government power with the states rather than the federal government.[8] This means that, especially if a state-run Medicaid program makes its connection to Medicaid less explicit, people may have a higher opinion of it.

In this study, I examine public perceptions of about Medicare, Medicaid, and state-based programs by analyzing sentiments expressed on Twitter. Social media platforms like Twitter provide a new universe of political expression to supplement public opinion polling and anecdotal evidence. By analyzing and aggregating Twitter data, I can gain a perspective on how politically engaged Twitter users frame these programs, especially in comparison to one another.

Questions and Hypotheses

---

[6] Hamel, et al.

[7] Cole, R., & Kincaid, J. (2000). Public Opinion and American Federalism: Perspectives on Taxes, Spending, and Trust: An ACIR Update. *Publius,* 30(1), 189-201. Retrieved from http://www.jstor.org/stable/3331128

[8] Gallup Poll, 2016, "Majority in U.S. Prefer State Over Federal Government Power," June 14-23, 2016, retrieved from https://news.gallup.com/poll/193595/majority-prefer-state-federal-government-power.aspx.

There are two main questions I will explore in this study. First, what words do Twitter users associate with government health care programs, and how do they compare to one another? I hypothesize that tweets about state-run Medicaid programs will use different and more specific keywords than tweets about the national Medicaid system. Because the connection between state-run programs and the national system may not be well-known, I think that tweets about the state-based programs would be much more localized and focus on specific political issues in each state rather than the larger political framing of national health care policy. This could mean that tweets about national Medicaid could mention "Trump" or "2020," while tweets about state-based Medicaid could mention governors' names or the name of the state capital. I also hypothesize that tweets about Medicare will focus more on Medicare as a stand-in for health care policy overall, while tweets about Medicaid will focus more on the specific political issue of Medicaid expansion. Especially because of the 2020 Democratic presidential primary, Medicare has become a focus of health care policy reform, specifically as a policy that should be sized up to include more people. This means that tweets about Medicare may function as tweets about the American health care system overall. Meanwhile, Medicaid has gotten comparatively less attention, so tweets about Medicaid may focus on the specifics of the program, like political movements around Medicaid expansion. In tweets about Medicare, I would expect to see keywords like "people" and "doctor," while in tweets about Medicaid I would expect to see keywords like "expansion" and "state."

My second question is: what sentiments (positive, neutral, or negative) do Twitter users express about these government programs? I hypothesize that, contrary to conventional wisdom, tweets about Medicaid will be more positive than tweets about Medicare. As stated in my prior hypothesis, I argue that Medicare as a concept has become

a stand-in for larger healthcare policy reform, so mentions of Medicare may have become more controversial and thus more negative. I posit that Medicaid, on the other hand, has undergone little political change in recent history, so may not have become as politicized as mentions of Medicare. I also hypothesize that tweets about state-based Medicaid programs will be more positive that tweets about the national system. As researchers have found, state governments have higher approval ratings than the federal government. This is likely to have an effect on perceptions of dual-administered programs and would make, for example, mentioning TennCare more popular than mentioning Medicaid.

Sample: Selection of State-Based Programs

Because each state administers their own Medicaid program, and often names the program differently, there are potentially fifty options for state-based programs to examine.[9] Some are easy to eliminate; many states, including New York, Texas, and Wisconsin, simply keep the name Medicaid, so would not be a useful comparison of the words themselves. Others, including California, Georgia, and Illinois, use vague names that could be used in reference to more general concepts, not the programs specifically, again making it difficult to discern the perceptions of the programs themselves. [10] In search of programs that are easily differentiable from the national Medicaid system and from general words about health care, for this study I will analyze Maine's program, MaineCare, and Tennessee's program, TennCare.

Maine is a particularly interesting case because of recent changes to its Medicaid program. In 2017, Maine voted in favor of a referendum to expand access to Medicaid

---

[9] "Homelessness Data in HHS Mainstream Programs. Medicaid Program Names," Office of the Assistant Secretary for Planning and Evaluation, *U.S. Department of Health & Human Services*, published 4 Feb. 2009, accessed from https://aspe.hhs.gov/dataset/medicaid-program-names.

[10] California's program is called "Medi-Cal," which is too similar to the adjective "medical" to be useful. Georgia, Illinois, and others refer to their programs as "Medicaid Assistance," which is a noun used independent of the programs.

under the Affordable Care Act.[11] Governor Paul LePage (R) resisted implementation of the referendum for the remainder of his term; in January 2019, he left office and was succeeded by Democrat Janet Mills, who signed an executive order implementing the referendum during her first week in office. This extended coverage under MaineCare to approximately 70,000 low-income people in Maine.[12] Under its newly Democratic state government, Maine also passed a law requiring all insurance and Medicaid to cover abortion procedures.[13] Because the national Medicaid system has remained fairly consistent, Maine is an interesting case to study the effects of significant health care reform on public perceptions, especially in comparison to the national dataset.

Tennessee offers an interesting complement to Maine's recent Medicaid history; while Maine's is a history of recent reform, Tennessee's is a history of recent problems with little movement on reform. Since 2017, hundreds of thousands of children have lost coverage under TennCare, usually because their families were no longer eligible or they failed to renew their eligibility.[14] Additionally, in 2019 Governor Bill Lee (R) submitted a proposal to the federal government to trade uncapped Medicaid funding for a block grant, giving the state government the freedom to determine TennCare eligibility.[15] Both of these

---

[11] Matthew Bloch and Jasmine Lee, "Election Results: Maine Medicaid Expansion," *The New York Times*, updated 20 Dec. 2017, accessed from https://www.nytimes.com/elections/results/maine-ballot-measure-medicaid-expansion.

[12] "Governor Mills Announces Federal Approval of Medicaid Expansion," State of Maine, released 3 April 2019, accessed from https://www.maine.gov/governor/mills/news/governor-mills-announces-federal-approval-medicaid-expansion-2019-04-03.

[13] Patty Wight, "Newly Blue, Maine Expands Access to Abortion," *NPR All Things Considered*, published 2 July 2019, accessed from https://www.npr.org/sections/health-shots/2019/07/02/737046658/newly-blue-maine-expands-access-to-abortion.

[14] Brett Kelman and Mike Reicher, "At least 220,000 Tennessee kids faced loss of health insurance due to lacking paperwork," *The Nashville Tennessean*, published 14 July 2019, accessed from https://www.tennessean.com/story/news/investigations/2019/07/14/tenncare-coverkids-medicaid-children-application-insurance-denied/1387769001/.

[15] Anita Wadhwani, "Federal government opens 'comment period' for controversial TennCare overhaul," *The Nashville Tennessean*, published 29 Nov. 2019, accessed from https://www.tennessean.com/story/news/investigations/2019/07/14/tenncare-coverkids-medicaid-children-application-insurance-denied/1387769001/.

events dominated the health politics of Tennessee in 2019; however, they are weaker focusing events than those in Maine. This means that TennCare, in the absence of the type of large reform that occurred in Maine, may have been influenced by national perceptions more, or at least provided a more static state to compare to Maine and national Medicaid.

Sample: Twitter

To examine public perceptions of government-run health care programs, I chose to analyze data from Twitter. Twitter, like most modern social media, offers a free method of political expression accessible to all Americans with Internet access, and is not restricted by the resources of a polling firm or research institution. Further, Twitter itself provides unique benefits as the source of our dataset in comparison to other social media sites. First, Twitter's continuous timeline, search functionality, and centralized source of content makes it easier to collect usable data; in comparison, a media site like Facebook has pages, groups, and profiles and generally lacks a public, stream-of-consciousness source of content. Second, Twitter is widely popular, especially for political engagement on social media. In 2019, 22 percent Americans reported using Twitter, making it one of the 10 most used social media platform in the United States. [16] Further, Twitter is widely regarded as the platform that draws the most political engagement and dissemination, with 31 percent of all users somehow engaging in a political conversation on the site. [17]

That being said, Twitter is an imperfect model for the general public. A study from Pew confirmed the refrain "Twitter is not real life;" 80 percent of all tweets come from 10 percent of users, and Twitter users are younger, more highly educated, wealthier, and more

---

[16] Pew Research Center, "Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018," published 10 Apr. 2019, retrieved from https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

[17] Pew Research Center, "A small group of prolific users account for a majority of political tweets sent by U.S. adults," published 23 Oct. 2019, retrieved from https://www.pewresearch.org/fact-tank/2019/10/23/a-small-group-of-prolific-users-account-for-a-majority-of-political-tweets-sent-by-u-s-adults/

likely to be Democrats and liberal than the general public. [18] These limitations have to be kept in mind throughout the course of this study and potential application of its results.

Although Twitter may not be a mirror of Americana, it does play a deeply important role in political discourse that should not be ignored. In a *New York Times* piece investigating the power dynamics of Senator Kamala Harris' (D-CA) 2020 presidential campaign, the reporters cited "the fixation some younger staffers [had] with liberals on Twitter" as something that "distorted their view of what issues and moments truly mattered"; [19] five days after the article's publication, Harris dropped out of the race. While Twitter may not always reflect the general public, that doesn't change the impact it can have on those most active on it, especially politicians that treat it as an accessible public forum; this means that Twitter is a uniquely politicized, if imperfectly representative, platform for understanding political expression.

Sample: Choosing the Dataset

As will be explained later, my method of data collection allows for collection as early as 2006. However, a dataset spanning 13 years is not advisable for my analysis of aggregated data, as it would make data collection, management, and analysis overwhelming.

I chose to limit my dataset to tweets published between February 1, 2019 and December 1, 2019. This provides a sizable dataset of ten months, yet is more compact than thirteen years. There was not a news-making nationwide healthcare event in this window, so the data is unlikely to be impacted by focusing events, unlike if, for example, the analysis

---

[18] Pew Research Center, "Sizing Up Twitter Users," published 24 Apr. 2019, retrieved from https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

[19] Jonathan Martin, Astead W. Herndon, and Alexander Burns, "How Kamala Harris's Campaign Unraveled," *The New York Times*, published 29 Nov. 2019, retrieved from https://www.nytimes.com/2019/11/29/us/politics/kamala-harris-2020.html

was done during the repeal and replace debates. This allows me to examine public opinion more generally, as focusing events drive opinion formation and bring out partisanship. February 1 makes the dataset less susceptible to the effects of mobilization after the 2018 elections and the minor focusing event of officials taking office. [20]

I will admit to a few limitations of this selected dataset, specifically the impact of the 2020 presidential campaign. While the first presidential primary or caucus is not until February 3, 2020, the campaign is well underway. Health care has become the focal point of the 2020 Democratic presidential primary, rivaled only by who is best equipped to win; in the first three debates, all within my dataset, candidates discussed health care for 90 minutes, twice as much as all foreign policy combined. [21] In concert, candidates have proposed numerous health care policy solutions, often dealing with Medicare and Medicaid. This could create problems in our dataset, as the program may pick up mentions of plans. I also cannot filter out the more subconscious effect that such widespread, Democrat-focused health care debate will have on the public perception of existing government programs.

Methodology

To collect tweets for analysis I used an open source script found on Github called *twitterscraper*. Twitter restricts data collection in its own interface by requiring users to apply through the Twitter Developers Lab and then limiting data collection to 3200 tweets from the prior three months. Because I would prefer greater free reign of data collection, I chose to use *twitterscraper*, which bypasses the Twitter Developer interface and allows for

---

[20] This does not account for special elections or the 2019 general elections as focusing events. While this may have an effect, I believe that the effect would be less substantial than during and after the 2018 elections, although admittedly not negligible.

[21] Hannah Brown and Dylan Scott, "The Democratic debates have spent 93 minutes on health care," *Vox.com*, published 15 Oct. 2019, retrieved from https://www.vox.com/policy-and-politics/2019/10/15/20914415/democratic-debates-health-care-issues

infinite data collection. While not the most formal methodology, for an undergraduate paper, I believe that I can stretch the limits of what might be okay to do.

twitterscraper collects a large amount of information about each tweet including username, text of the tweet, timestamp, number of likes, retweets, and replies. In this study, I will only utilize the texts of the tweets and the usernames, although the other components make future studies possible. An explanation of the code I used is attached in Appendix A.

I use three mechanisms for aggregating tweets and creating a useful and holistic analysis: (1) word frequency, (2) sentiment analysis, and (3) sentiment analysis aggregated by user. Analysis of word frequency explores the words users associate with and use alongside mentions of these programs. Especially when filler words, like articles and pronouns, are removed, this can provide important and interesting insight into what things are most-commonly associated with these programs and how users perceive them relative to other events and issues. Sentiment analysis produces the useful analysis of whether tweets are more positive, more neutral, or more negative in their point of view. This provides an understanding of the attitudes and feelings behind Twitter conversations about these programs. Initial analysis of sentiment examines tweets, not users; this means that, if one user tweets frequently and negatively about a program, they will be overrepresented in the dataset. Thus, a sentiment analysis of individual users is also interesting and useful to better understand what feelings are over-represented in tweets relative to active Twitter users.

Results

twitterscraper produced a random dataset of tweets from each keyword with a limit set at 5040. It collected 5040 tweets mentioning “medicaid,” 5040 mentioning “medicare,” 2944 mentioning “tenncare,” and 810 mentioning “mainecare.” Because tweets about

Medicaid and Medicare reached the tweet limit, this is not necessarily a proportional sample.

Word Frequency

The analysis of word frequencies produced four datasets of words and number of appearances in the tweets for each keyword. The Medicaid dataset used 15,158 unique words, the Medicare dataset used 13,080 unique words, the TennCare dataset used 9,059 unique words, and the MaineCare dataset used 3,402 unique words. Because filler words like "the" are by far the most common but also offer very little information, I manually removed a list of 73 filler words that consistently appeared in the top 100 words of the datasets; most of these words were articles and pronouns.

The following table shows the most-common words in tweets mentioning "medicaid." Logically, the most-common word is "medicaid." Beyond that, this dataset shows a dual focus on the mechanics of health care and Medicaid, through words like "insurance," "states," and "requirements," and on political reforms through words like "trump," "expansion," and "cut." This dataset also displays a general knowledge of the relationship between Medicaid and employment, as "work" is the fourth most-common word. Interestingly, "social" is the eighth most-common word; especially because "security" also appears in this dataset, it is possible that this reflects a large number of references to Social Security.

Most-Common Words for Tweets Mentioning "Medicaid"

medicaid	we	care	more	insurance
medicare	health	trump	my	states
people	all	expansion	healthcare	requirements

work	social	security	pay	cut

The following table shows the most-common words in tweets mentioning "medicare." Again, the most-common word is, logically, "medicare." This dataset is more general than the Medicaid dataset, as many words refer more generally to health care concepts like "people," "care," and "plan." There is also explicit mention of Senator Bernie Sanders (D-VT), the lead proponent of "Medicare for All" and a 2020 presidential candidate. Like in the Medicaid dataset, there are mentions of "social" and "security," likely in reference to Social Security. Interestingly, this dataset explicitly mentions "private," presumably referring to private insurance; it's possible that this is in response to Sanders' "Medicare for All" plan to essentially eliminate private health insurance. Overall, this dataset is likely heavily influenced by "Medicare for All," as evidenced by the prominence of "all," "bernie," "sanders," and "private."

Most-Common Words for Tweets Mentioning "Medicare"

medicare	people	health	security	more
all	care	plan	bernie	medicaid
we	healthcare	social	sanders	us
insurance	pay	my	want	private

The next table shows the most-common words in tweets mentioning "mainecare." Many of the words are related to the health policy events in Maine in 2019. Words related to time, such as "todays" and "calendar," are likely related to the restricted and delayed timeline of MaineCare expansion. The prominence of "abortion" and "abortions" is likely a result of abortion service expansion under MaineCare in 2019. There is an interesting

dichotomy between personal and political experiences in this dataset. The prominence of "id" could be evidence of more personal tweets about health care policy, as users may be referring to their own MaineCare ID card.[22] Yet, the prominence of "bill," "sp," and "hp" are evidence of more political topics in the dataset, as "sp" and "hp" likely refer to the codes for Maine Senate and House legislation. It should also be noted that, because I filtered out the word "me," I also filtered out references to Maine's state acronym ME, likely influencing the dataset.

Most-Common Words in Tweets Mentioning MaineCare

mainecare	bill	services	my	more
maine	id	todays	insurance	sp
coverage	health	calendar	abortions	hp
abortion	care	cover	provide	people

The next table shows the most-common words in tweets mentioning "tenncare." This dataset is very specific, even beyond its focus on an individual state; unlike the Medicaid and Medicaid datasets, which frequently refer to general concepts like "people" and "healthcare," the TennCare dataset focuses further on specific aspects of healthcare, like "block" "grant[s]", and specific populations, like "kids" and "children." This is likely reflective of the Medicaid block grant proposal in Tennessee that then acted as a focusing event for TennCare. This dataset also makes mention of "fraud," an interesting and unique mention not reflected in the other datasets.

Most-Common Words for Tweets Mentioning "TennCare"

[22] Maine also has a recent history of debates over voter ID laws, so there is a minimal possibility that this is in reference to those political debates.

tenncare	medicaid	children	proposal	tn
block	state	insurance	we	care
grant	health	more	my	via
tennessee	kids	coverage	fraud	program

The following table shows the twenty nine words that constitute the top ten words from each dataset and their relative rank in each dataset. Based on this table, the Medicaid and Medicare datasets are fairly dissimilar. Significantly, “work” and “trump” are the fourth and tenth most common words mentioned in tweets about Medicaid, while they are below or nearly below the fiftieth rank in tweets about Medicare. Additionally, tweets about Medicaid frequently mention “medicare,” making it the second most common word; yet, tweets about Medicare mention “medicaid” significantly less at the twenty-second rank. Tweets about Medicare mention “insurance,” “healthcare,” and “pay” at very high rates, while tweets about Medicaid mention them below the fiftieth, at the twenty-third, and at the twenty-fourth ranks, respectively. Both datasets mention “people,” “we,” “health,” “all,” and “social” at high ranks.

The MaineCare and TennCare datasets are even more dissimilar from the Medicaid and Medicare datasets. “tenncare,” “block,” “grant,” “tennessee,” “state,” “kids,” and “children” are some of the most common words mentioned in tweets about Tenncare, yet they are all below the fiftieth rank for the Medicaid and Medicare datasets. MaineCare follows a similar pattern, as “mainecare,” “abortion,” and “id” are very common in the MaineCare dataset but very uncommon in the others. “insurance,” “health,” and “medicaid” are the only words from the ten most common words in the TennCare dataset that appear comparatively frequently in the Medicaid or Medicare datasets. Even the most common

words in the Medicaid and Medicare datasets are not very common in the TennCare dataset; “medicare,” “social,” “trump,” and “pay” are all below the fiftieth rank for tweets about TennCare. “people,” “work,” “we,” “all,” “care,” “no,” and “healthcare” all appear in the fifty most common words for tweets about TennCare, but at significantly lower ranks than in tweets about Medicaid or Medicare. Similarly, MaineCare makes little mention of “medicaid” or “medicare” explicitly, and also uses words like “work,” “we,” and “all” less.

Most-Common Words Comparison (Top 10 words from each dataset)

Word	“Medicaid” Rank	“Medicare” Rank	“TennCare” Rank	“MaineCare” Rank
medicaid	1	22	5	24
medicare	2	1	77	63
people	3	5	22	20
work	4	79	38	82
we	5	3	14	33
health	6	10	7	7
all	7	2	33	54
social	8	12	536	250
care	9	7	18	8
trump	10	49	258	N/A
insurance	17	4	10	14
state	20	271	6	21
healthcare	23	8	49	65
pay	24	9	78	22

coverage	28	47	12	3
services	56	145	74	9
bill	105	58	22	5
children	124	487	9	303
kids	161	495	8	521
block	536	1308	2	N/A
id	554	447	930	6
abortion	795	866	1555	4
grant	843	3635	3	1860
todays	1192	2419	836	10
tennessee	1432	N/A	4	N/A
maine	2905	5914	N/A	2
tenncare	3855	N/A	1	N/A
no	N/A	6	40	78
mainecare	N/A	N/A	N/A	1

Sentiment Analysis

This study's sentiment analyses were produced using a library that analyzes film reviews and rates them as positive, negative, or neutral. Because film reviews are different from tweets and politics, it is likely that this is an imperfect assessment of the sentiment of tweets. However, the usability and accessibility of this library is unmatched, and hopefully it will still provide a useful comparison.

The following table displays the distribution of positive, neutral, and negative assessments of tweets for each dataset. Based on these results, a plurality of tweets about

Medicaid, Medicare, and TennCare are positive. In the case of MaineCare, a plurality of tweets are neutral. All four programs have very high proportion of neutral tweets, potentially displaying either the imperfection of applying film review sentiments to political tweets or the prominence of impartial news tweets about these programs. Tweets about Medicaid have the highest negative proportion and tweets about MaineCare have the lowest negative proportion.

Sentiment of Tweets about each Program

	positive	%	neutral	%	negative	%
Medicaid	2355	*46.7*	1416	*28.1*	1269	*25.2*
Medicare	2463	*48.9*	1431	*28.3*	1145	*22.7*
MaineCare	309	*38.2*	387	*47.8*	114	*14.1*
TennCare	1234	*41.9*	1102	*37.4*	608	*20.7*

Sentiment Analysis of Users

While the above analysis is instructive, there are some further questions to ask about the results. The above sentiment analysis is an analysis of tweets rather than users; thus, if one user tweets multiple times about one of the programs, their (likely repetitive) sentiment is expressed multiple times in the dataset. This is not a great method of determining overall perception of these programs by users. To resolve this issue, I aggregated the sentiments of the tweets of each user and then rounded the aggregate to the nearest integer. The distribution of users, rather than tweets, is displayed in the table below. These results indicate that when discussing Medicaid and Medicare, the average user is more neutral than the average tweet. This means that there are some users tweeting multiple times positively or negatively about Medicaid or Medicare, impacting the original

dataset. The MaineCare and TennCare datasets exhibits the opposite phenomena, as the average user is more positive, less neutral and less negative than the average tweet. This means that the dataset of tweets is not representative of the opinions of the Twitter population, much less the general population, as positive and negative tweets are overrepresented in the Medicaid and Medicare datasets and negative tweets are overrepresented in the MaineCare and TennCare datasets.

Sentiment of Users who Tweet about each Program

	positive	%	neutral	%	negative	%
Medicaid	1765	*43.7*	1320	32.7	956	*23.7*
Medicare	1892	*45.5*	1412	34.0	854	*20.5*
MaineCare	187	*50.5*	134	36.2	49	*13.2*
TennCare	560	*46.6*	416	34.6	227	*18.9*

Synthesis

My results indicate significant differences between the Twitter conversations about Medicaid, Medicare, MaineCare, and TennCare.

First, the number of tweets varies greatly by program. While my samples are of the same size, during data collection I discovered that the available dataset for Medicare was significantly larger than the available dataset for Medicaid. I attribute this to the far-reaching conversation about Senator Sanders' "Medicare for All" proposal and the relative lack of focusing events around the national Medicaid program. For state-based Medicaid programs, the available dataset for MaineCare was much smaller than the available dataset for TennCare. This is likely simply attributable to Tennessee's larger population, although

there could be some other impacting factors related to the states' specific demographic and political environments.

My second conclusion is that, despite being in the same policy area, very different words are used in association with each program. While the top ten words for each program have some overlap, they are mostly unique. Even though Medicare and Medicaid are both administered by the federal government and many people associate or confuse them, they are actually not talked about using the same keywords. Particularly notable is the prominence of "work" in tweets about Medicaid, compared to "work" as only the 79th most common word concerning Medicare. This is likely related to Medicaid work requirements, a contemporary debate in the politics of Medicaid.[23] Also interesting, and more surprising, is the relative number of references to President Donald Trump (R); his name is mentioned very frequently in tweets about Medicaid, but less frequently in tweets about Medicare. I argue that this may be an effect of the 2020 Democratic presidential primary; it is possible that most of the Medicare dataset encompasses tweets about the Democratic primary process, an idea supported by the prominence of Bernie Sanders in the dataset. Thus, the Medicare dataset may be focused on a Democratic presidency and future policies independent of Trump, while the Medicaid dataset may be focused on contemporary policies independent of the Democratic presidential candidates.

Another important conclusion of the word frequency analysis is the relative independence of the MaineCare and TennCare datasets. While officially Medicaid programs, the Twitter conversations about these programs don't use the same keywords as conversations about the national program. Instead, the TennCare and MaineCare datasets

---

[23] Abby Goodnough, "South Carolina is the 10th State to Impose Medicaid Work Requirements," *The New York Times*, published 12 Dec. 2019, accessed from https://www.nytimes.com/2019/12/12/health/medicaid-work-requirements-SC.html.

contain keywords nearly unique to their programs and political environments and events, such as “block grants,” “abortion,” and “kids.” This provides important insight into the relationship between perceptions of state-based and the national Medicaid programs.

I can see two possible reasons for this outcome. First, perhaps people treat Medicaid as a national program as a separate entity from the state-based programs, and think of programs before issues. Because people view them as independent programs, when people consider Medicaid, they think of issues like Medicaid work requirements and national partisan action from President Trump. Further, when people consider state-based Medicaid programs, they think not of national issues, but of state-specific issues like abortion coverage in Maine or disenrollment in Tennessee. Conversely, maybe people think of issues and then associate them with specific programs. Issues like Medicaid work requirements or action from President Trump are seen as impacting Medicaid, but not the state-based programs that administer them. Further, more general health care issues are seen as impacting Medicare rather than other government programs because of the prominence of Medicare in current national debate.

My final conclusion is that sentiment analysis of tweets does not expose many explicable or important differences, but comparison of tweets and users does. Overall, the sentiments expressed in tweets about the four programs seem to be fairly comparable, although somewhat variable. A big factor is the large number of neutral tweets; it’s unclear if this is a result of many impartial news reports or the algorithm struggling to identify a sentiment in many tweets. The results of the tweet sentiment analysis are not very similar to the public polling that has existed; while this could be a new discovery of this study, more likely it is an error in the sentiment analysis.

More interesting, however, is the difference between the distribution of tweets sentiments and the distribution of the average sentiment of each user. The average user is more negative and less positive about Medicaid and Medicare than the average tweet, and the average user is less negative and more positive about MaineCare and TennCare. This is the result of my study that most opens doors for future research. This implies that users with positive perceptions of Medicaid and Medicare talk about the programs more on Twitter, but the opposite is true for MaineCare and TennCare, as users with negative perceptions talk about those programs more on Twitter than users with positive perceptions. This result provides some evidence that MaineCare and TennCare are more popular than national Medicaid, even if their users are less vocal about it, as this is proof that users negative about MaineCare and TennCare and users positive about Medicaid and Medicare take up more of the conversation, potentially disguising the true distribution of beliefs. This needs more exploration in a more thorough study, likely branching beyond the confines of this specific comparison, including these four programs as well as other policies in Maine and Tennessee and other states' health care systems.

Appendix A. Explanation of Code Methods

My code is written in Python 3, one of the most common programming languages, and the files are stored in the Jupyter Notebook format, a common web application used to execute programming languages.

twitterscraper uses the structure of Twitter search queries to collect data. For example, to collect data on Medicaid mentions, I enter "Medicaid until:2019-12-01 since:2019-02-01" into the *twitterscraper* structure. When collecting data, I store the list of collected tweets in a *.json* file for easy access.

Each tweet is stored in a data format called a "dictionary." A Python dictionary assigns an element, like a number, phrase, or list of things, to a "key," usually a word or phrase. For example, each tweet dictionary has a key, "likes," which is connected to an integer, for example "27;" this means that this tweet had 27 likes at time of collection. Each tweet dictionary contains 16 different keys, storing items such as the name and username of the user, the text of the tweet, the timestamp, and the number of likes, retweets, and replies.

Python's interface is sensitive to case and punctuation; to make the data most usable, I used iterative code to make all letters in each tweet lower case, and used an advanced package called *regular expressions* to remove all punctuation.

In general, my mechanisms iterate over the list of tweets to collect, calculate, and then sort data. To assess the frequency of certain words, I create one long phrase, called a "string" and then use a built-in function to count the frequency of each unique word in the string. To assess the most active users, I use a built-in function, "set," to create a list of unique users and then iterate over that list, creating dictionaries that contain, for the most active users, the username, number of tweets, and the texts of those tweets.

For sentiment analysis, I use a library called *textblob*. *textblob* is a library that executes a variety of textual processes in Python, including sentiment analysis. To determine sentiment

analysis I create a new column in my dataset and then iterate through each of my datasets, looking at each tweet text and using *textblob* to determine the sentiment as an integer: -1 (negative), 0 (neutral), and 1 (positive). For the user sentiment analysis, I group the dataset by username and then calculate the mean of the sentiments, rounding it to the nearest integer to get the average sentiment of each user.

Appendix B. Top 30 Words for Each Program's Dataset

	Medicaid	Medicare	MaineCare	TennCare
1	medicaid	medicare	mainecare*	tenncare*
2	medicare	all	maine*	block*
3	people	we	coverage	grant*
4	work*	insurance	abortion*	tennessee*
5	we	people	bill	medicaid
6	health	care	id*	state
7	all	healthcare	health	health
8	social	pay	care	kids*
9	care	health	services*	children*
10	trump*	plan	todays*	insurance
11	expansion*	social	calendar*	more
12	security	my	cover*	coverage
13	more	security	my	proposal*
14	my	bernie*	insurance	we
15	healthcare	sanders*	abortions*	my
16	pay	want*	provide*	fraud
17	insurance	more	more	tn*
18	states*	medicaid	sp*	care
19	requirements*	us*	hp*	via*
20	state	private*	people	program
21	cut	our	state	people
22	need	free*	pay	bill
23	up	now	amend*	tennessean*
24	plan	up	medicaid	plan
25	programs*	think*	our	our
26	program	fraud	requiring*	after*
27	federal	only*	now	public*
28	coverage	need	require*	federal
29	tax	tax	act*	all
30	gop*	new*	mepolitics*	now

**unique words in this table*