WHAT AFFECTS ONE, AFFECTS ALL: UTILIZING TWEETS TO MEASURE THE
SENTIMENT OF NEW JERSEY COMMUNITIES APPROACHING THE “SNAP
GAP”

By
TORY RICHARD MASCUILLI

A thesis submitted to the
Graduate School-Camden
Rutgers, The State University of New Jersey
In partial fulfillment of the requirements
For the degree of Master of Arts
Graduate Program in Psychology
Written under the direction of
Dr. Daniel Hart
And approved by

______________________________
Dr. Daniel Hart

______________________________
Dr. Sarah Allred

______________________________
Dr. Robert Atkins

Camden, New Jersey
May 2021
THESIS ABSTRACT

What Affects One, Affects All: Utilizing Tweets to Measure the Sentiment of New Jersey Communities Approaching the “SNAP Gap”

by TORY RICHARD MASCUILLI

Thesis Director:
Dr. Daniel Hart

The Supplemental Nutrition Assistance Program (SNAP) provides monthly funding to low-income participants to buy food. Research has shown that this supplemental assistance is insufficient to cover an entire month’s needs, leading to a period of time referred to as the “SNAP gap,” during which benefits have been depleted. Food insecurity, amplified by the exhaustion of SNAP benefits, may increase negative emotions, which may be detected at the community level. The current study examined the possibility that as food insecurity increased over the course of a month as SNAP benefits were exhausted in areas heavily dependent on SNAP so too would negative affect measured at the level of the community. To test for this effect, tweets were collected from each of the 21 county seats of New Jersey, and coded for emotion words. Negative affect words peaked during the SNAP gap. The effect was more pronounced in communities with higher levels of poverty. Implications and directions for future research are discussed.
Dedication

In memory of my parents, Richard and Christine
Acknowledgements

Special thanks to Charlotte Markey for the encouragement in pursuing this degree.

Thank you to my committee members, Sarah Allred and Bob Atkins, for helping to guide me in the refinement of my thesis. I am grateful for your time and your wisdom.

I would like to thank my advisor, Dan Hart, whose mentorship extends far beyond this project. Let it be known: Stack Overflow ain’t got nothin’ on you!

Finally, I would like to thank Rosie, for the support, the unconditional love, and for believing, even when I didn’t. I love you.
Introduction

In 2020, the Supplemental Nutrition Assistance Program (SNAP) provided food-purchasing assistance for around 670,000 New Jersey residents, or approximately 340,000 households (Office of the Governor of New Jersey, 2020). In the time since the inception of the program, researchers have examined the broader impact of SNAP on a number of issues, among them the problem of food insecurity. Food insecurity is defined by the United States Department of Agriculture (USDA) as a lack of ability by an individual or group of individuals to obtain nutritious foods (Bickel et al., 2000).

Empirical evidence suggests that enrollment in SNAP reduces but does not eliminate food insecurity (Wilde, 2013). Hamelin et al. (1999) discuss the social and community implications of food insecurity, labelling it as a community stressor and as a contributor to the persistence of socioeconomic inequities. Food insecurity was also found to be associated with numerous deleterious effects on mental health and emotional well-being, such as high levels of negative affect (Bruening et al., 2017).

As recently as 2013, food insecurity in the United States impacted around 50 million individuals, and 95% of these were SNAP participants. (Gundersen & Ziliak, 2015). A large body of evidence has linked food insecurity to myriad negative health outcomes. An individual or family’s lack of ability to obtain nutritious food has been associated with poor mental health and has been found to co-occur in populations that measured high on levels of shame, stress, anxiety, aggression, suicidal thoughts, and depression (Pryor et al., 2016). The risks posed by food insecurity extend beyond the confines of the individual and family household and have broader social implications on health and well-being at the community level. Hamelin et al. (1999) discuss how the
psychological stress of chronic food insecurity within the household can lead to the developmental disruption of society and the sustentation of socioeconomic inequities; it is labelled as a threat to community harmony due to the intensification of negative emotional states. Research delineating the mechanisms by which this relationship occurs suggest that as levels of food insecurity increase, individuals, families, and ultimately, communities, are more likely to experience growing concern over a lack of food transform into one or several of the aforementioned negative emotional states (Bruening et al., 2017).

**SNAP, Food Insecurity, and Community Health Correlates**

The provision of SNAP benefits comes in the form of a Families First Electronic Benefits Transfer (EBT) debit account, with participants of the program receiving a deposit once per month, typically within the first two weeks of the month. In the state of New Jersey, benefits are dispersed within the first five days of the month and are calculated based on several factors, including household size and income (DHS, 2020). These deposits are estimated to be sufficient to cover the food-purchasing needs for that individual or family for the subsequent month. Recent data suggest that across all household sizes, SNAP funding remains insufficient to cover an individual or family’s needs in terms of food (U.S. Council of Economic Advisers, 2015). Additional research has shown that by extension, as benefits run out nearing the end of each month, food purchases and food consumption decrease in populations that rely on SNAP (Hoynes & Schanzenbach, 2015; Todd & Gregory, 2018). This end-of-month period in which participant funding has depleted has been referred to by Laurito and Schwartz (2019) as the “SNAP gap.” Decreases in both food purchasing and consumption during this gap in
the benefits cycle may be reflective of the food insecurity that many participants of SNAP struggle with.

**Negative Affect and Sentiment Analysis**

Negative emotional states such as stress, anxiety, and depression are associated with poor mental health outcomes when experienced chronically (Koch et al., 2013; Watson & Clark, 1984). In a seminal paper on the subject, Watson and Clark (1984) delineate a model suggesting that negative affect can be largely reactive in nature, the result of an individual’s situational environment. Building on the work of Watson and Clark, Berry and Hansen (1996) suggest that negative affect is measurable in individuals and may be categorized as high-level or low-level. High-negative affect and low-negative affect each consist of emotions of a similar valence. For instance, a person with high-negative affect will experience emotional states such as anxiety, depression, anger, and stress; these emotions may be experienced simultaneously or in isolation. It is the frequent exposure to these negative affective states that place a person at greater risk for poor social behavior and poor health outcomes (Mayne, 1999).

Measurement of a person’s display of affect may be performed through analysis of their writing (Batson et al., 1992). Researchers typically rely on lexical approaches to measuring affect in writing, utilizing specific types of words or language to indicate positive or negative affect (Osherenko & Andre, 2007). With access to increasingly large databases of written word, thanks in large part to the availability of social media data, lexical analysis can be done by computer.

Computer programming techniques have shown promise with regard to navigating largescale databases, such as those provided by social media websites (Kharde
& Sonawane, 2016). Text-based social media platforms, specifically Twitter, have become popular amongst psychology and public health researchers for exploring emotion-related phenomena in large, demographically diverse populations (Sinnenberg et al., 2017). Colditz et al. (2018) made significant progress in terms of navigating largescale social media databases for the purposes of public health and behavioral analysis, specifically within the Twitter database. The researchers were able to circumvent the previously mentioned obstacles by utilizing methods derived from machine learning. These methods included performing data mining on subsets of tweets and then applying a program to those tweets that would perform sentiment analysis.

Sentiment analysis is a process by which the sentiment, or affective tone, of a word, phrase, or sentence is analyzed and then categorized based on the emotional content of the written words (Agarwal et al., 2011). This process is one widely used approach to measuring a person’s affect through their writing and may be completed either by manual annotation or with computer programs. As previously mentioned, manual annotation of a database as large as a social media website such as Twitter would prove to be too laborious, thus the need for more sophisticated programming methods. Agarwal et al. (2011) utilize such computer programming models in their sentiment analysis of Twitter data, delineating a system of classification of tweets as positive, negative, or neutral in affect.

Given the lack of research monitoring changes in the affective state of populations at risk of food insecurity, the primary interest here was to track community sentiment longitudinally via Twitter. New Jersey communities approaching the SNAP gap were
hypothesized to tweet messages of an increasingly negative affective tone, reflecting the exhaustion of benefits.

The measurement of time in this study is the day of the month. Recall that the prediction is that the SNAP gap will cause the last days of the month and the first days of the month to be highest in negative emotion (these are times when SNAP benefits have been exhausted [end of the month] and in the days just prior to the arrival of the monthly benefits [the first days of the month]). This means that plotted against day of the month (1-30), negative emotion will start high, slowly diminish (days 7-21), then increase (days 22-30). This leads to the prediction of a quadratic relation between day of the month and negative emotion.

Finally, the SNAP gap is likely most pronounced in communities with high levels of poverty. This leads to the prediction that poverty level will be a predictor of the magnitude of this quadratic association.
Methods

Units of Analysis

County seats were the units of analysis in this study. Community sentiment in each county seat was inferred from the tweets of residents living within a 5-mile radius of the geographical center of each county seat. Because tweets are contributed by many individuals who cannot be identified, it is not possible to provide an accurate assessment of the sociodemographic characteristics of those whose tweets were included in the collection and final analysis. Instead, this project collected and analyzed tweets from communities with the goal of creating an aggregate sentiment score for the affective state of each of the 21 county seats of New Jersey for each day during the examination period.

Procedure

A script was written within the statistical environment R. For the tweet collection process, the Rtweet package was utilized. Rtweet allows users to write a script and then specify a number of variable parameters under which the data may be collected. These parameters include information such as specific keywords, desired number of tweets, and geographic location. Rtweet performs these operations through a process called web scraping, which is a method of quickly scanning largescale databases for the chosen parameters in an automated fashion. The script written for this project was programmed to gather approximately 5,000 original tweets from each of the 21 county seats of New Jersey during each collection interval. A filter was applied through the script to exclude the collection of any retweets, so as to avoid any duplicates during sentiment analysis. Geographic location was achieved by utilizing the latitude and longitude of each county seat.
Tweets were collected across four months— from March 2020 to June 2020— at a frequency of once per week, for a total of 16 collections.

Additional script was written in R to perform lexical analysis of the sentiment of each tweet, which were subsequently assigned a positive affective state, a negative affective state, or a neutral affective state. As this work was strictly observational in nature— that is, nonexperimental— no conditions were assigned, nor any interventions applied during the course of the project. Because this study focused on communities, not individuals, the nature of this project and the methodology being employed qualified it as exempt under the Institutional Review Board (IRB), as it was strictly observational in nature.

Measures

**Independent Variables.** The independent variables of this study were the day of month and the community’s poverty level. Data from the U.S Census was used to characterize communities’ poverty levels.

**Dependent Variable.** The dependent variable of this study was the average negative sentiment of all the tweets collected in a county seat for each day during the collection period.

**Sentiment and Data Analyses**

Both sentiment and data analyses also relied upon R. Sentiment analysis was performed through the use of the R package quanteda. Quanteda allows for the management and quantitative analysis of text corpora. An additional function of the quanteda package is the ability to input external dictionaries for the purposes of sentiment analysis; here, the dictionary of reference utilized was the Linguistic Inquiry and Word
Count (LIWC), created by Pennebaker et al. (2003). LIWC is widely used as a reference in providing psychological insight into written text; the dictionary contains approximately 4,500 words and word stems organized by various categories, including psychological. Imported texts may be compared word-by-word to the features of these categories, allowing for the identification of sentiment. Sentiment scores are assigned a valence of positive, negative, or neutral, as well as an intensity based on the inherent strength of the words and the context in which those words are used. Here, each individual tweet collected was assigned a negative emotion score; these scores were then aggregated at the level of each county seat, for each day.

The initial step in the data analysis process was to isolate the variables of interest. The LIWC database includes a category for negative emotions. This category of words, or dictionary, includes words that carry negative sentiment (e.g., agony, despair, fear, and harm). Each tweet was coded for the number of words it included that matched those in the negative emotion dictionary, with this sum divided by the number of words in the tweet and then multiplied by 1000.

Data from the U.S. Census (Census, 2018) estimating the percentage of the population living below the poverty line was used to characterize levels of poverty in each county seat. The hypothesis was then tested using a linear mixed-effects model. This approach was chosen to account for the nested structure of the data: the first level was the day of the month, the second level was each individual tweet, and the third level was the county seat each tweet was collected from. The model tested whether there was an interaction effect from the level of poverty in each county seat by day of the month that predicted negative sentiment levels.
Results

Table 1 presents descriptive statistics for the negemo variable and for poverty level:

Table 1

*Descriptive Statistics for negemo and Poverty Level*

<table>
<thead>
<tr>
<th>Measures</th>
<th>$M$</th>
<th>$SD$</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>negemo</td>
<td>2.15</td>
<td>5.31</td>
<td>0 - 100</td>
</tr>
<tr>
<td>Poverty Level</td>
<td>20.80</td>
<td>10.00</td>
<td>6.0 - 41.3</td>
</tr>
</tbody>
</table>

The hypothesis predicted an increase in the negative sentiment of communities approaching the “SNAP gap,” with the effect expected to be pronounced in communities with higher levels on average of poverty.

Figure 1 plots a representation of the average negative sentiment of each county seat across an average of the days of the collection period, grouped by average poverty level.

The figure suggests that the predicted associations exist.
Figure 1. Line plot with loess-smoothed lines of average negative sentiment of county seats as a function of the day of the month and average poverty level. Poverty level represents the percentage of the population living below the poverty line in each county.
Table 2 presents the results of the statistical test of the hypotheses.

Table 2

*Fixed Effects of Key Variables and Interaction Term*

| Measures              | Estimate | SE    | df      | t value | Pr(>|t|)   |
|-----------------------|----------|-------|---------|---------|------------|
| (Intercept)           | 2.032    | 0.129 | 19.67   | 15.716  | < 0.0000 *** |
| Day                   | -0.021   | 0.003 | 1052000 | -7.537  | < 0.0000 *** |
| Poverty Level         | 0.012    | 0.006 | 19.23   | 2.192   | 0.0409 *   |
| I(Day^2)              | 0.001    | 0.0001| 1052000 | 7.19    | < 0.0000 *** |
| Day:Poverty Level     | 0.0001   | 0.0001| 1052000 | 2.104   | 0.0354 *   |

*Note:*** p < 0.001. * p < 0.05.*

The results presented in Table 2 suggest that the hypothesis was confirmed. Indeed, negative sentiment was found to increase during the hypothesized period of time - that is, closest to the 20th of each month. The effect was found to be especially pronounced in less affluent communities, as predicted.
Discussion

To reiterate, the goal of this study was to monitor the sentiment of New Jersey communities via Twitter, with increases in negative affective tone expected approaching the “SNAP gap.” These increases were expected to be particularly pronounced in less affluent communities that may be more heavily reliant on SNAP benefits. To monitor sentiment, a script written in R collected tweets over the course of four months; additional script performed sentiment analysis on these tweets, relying upon the LIWC dictionary first designed by Pennebaker et al. (2003). The analysis suggests that negative affective tone increases approaching the final days of each month. Moreover, the findings indicate a statistically significant interaction between the day of the month and the level of poverty in each community, suggesting that tweets in poor communities have higher levels of negative affect surrounding the “SNAP gap.”

Based on the analysis, it would appear that the hypothesized effect is strongest with regards to the day of the month. Negative affect gradually increases over the course of each month, peaking during the expected period. The affective tone then gradually shifts to a more positive valence closer to the start of the next month. As noted, level of poverty appears to play a role in the general affective tone of each community, such that the higher the average poverty rate, the higher the level of negative affect. This finding is consistent with previous research that suggests that poverty plays a role in increasing levels of negative affect (Adamkovič & Martončík, 2017; Haushofer & Fehr, 2014). It is safe to imagine that impoverished communities would be generating tweets with greater negative sentiment when considering the psychological effects of poverty; when
combined with the potential exhaustion of SNAP benefits, it follows that there would be a strong interaction in the data.

It is worth noting, as depicted visually in Figure 1, that the average negative emotion of communities changes following the “SNAP gap.” This may reflect a decrease in the usage of words that carry a negative valence, an increase in the use of words that carry a positive valence, or a combination of the two. While the pattern broadly fits that which was predicted, the positive shift in sentiment occurred sooner than expected. One possible explanation for this may be that as the start of a new month approaches, less negative words are being utilized and sentiment may be more anticipatory in nature, potentially due to upcoming renewal of benefits.

**Limitations**

One limitation of this study is that it may not be possible to directly attribute the effect to exhaustion of benefits nearing the “SNAP gap.” While negative sentiment increased and decreased in a pattern relatively similar to the current study’s prediction, more data would be required to ascertain if the effect follows the same pattern consistently over extended periods of time. It was also not possible to ensure with any degree of certainty that tweets were being collected directly from communities that rely on SNAP; statistically, it seems plausible that a portion of the population utilize these benefits when considering the large number of New Jersey residents enrolled in the program (Office of the Governor of New Jersey, 2020). It is possible that other monthly benefits programs, such as the Temporary Assistance for Needy Families (TANF) program, contribute to this cyclical pattern of increases and decreases in affective tone.
Twitter data is useful for gleaning insights into the public’s reaction to any number of events (Thelwall, 2014), a detail that is particularly salient to the current study. During the data collection period, the COVID-19 pandemic and the murder of George Floyd- which lead to subsequent worldwide protests- were actively occurring. This is important to note as it is likely that there were numerous and consistent reactions to these events on Twitter, which potentially created a high level of noise within the data.

Implications

One potential implication of this study is that it may show an alternative form of evidence for the existence of the “SNAP gap.” While the study does contain the previously mentioned limitations, it is still of interest that the effect appears to be strongest during the hypothesized period of time and follows a pattern that is fairly consistent with the dispersal of benefits. In light of this possibility, it is important to note that this study also contributes to prior research showing that SNAP benefits are insufficient to cover the needs of those enrolled in the program. If communities, and particularly high poverty communities, are experiencing food insecurity despite enrollment in SNAP, that would suggest that a reevaluation of the program might be necessary, up to and including an elevation in the benefits rate sufficient to meet participant needs.

This study also contributes, in part, to the discourse on the risks of chronically high levels of negative affect. If indeed residents of New Jersey communities are experiencing food insecurity as a result of depleted SNAP benefits, then the findings would be consistent with the work of Hamelin et al. (1999), in that negative affect can
extend beyond the individual and family household and be observed at the community level.

**Future Directions**

Future research would benefit from a dataset that is collected over a significantly longer period of time than the one analyzed in this study. This would lend some credence to the possibility that these patterns of increases and decreases in affective states do indeed follow that of the benefits cycle.

As noted, the confluence of the COVID-19 pandemic and the worldwide reactions to the murder of George Floyd likely influenced the data used in the analysis. Future research that relies upon big data measures capable of parsing out the content of tweets might be more efficacious in approximating the type of change predicted by this study.

Finally, longitudinal research following individuals and families from communities of differing levels of poverty who are also reliant upon SNAP benefits might prove to be more sensitive to the effects detected in this study.

In general, there is a need for more research examining the consequences of negative affect in poor and food insecure communities. Examining real time reactions from resources such as Twitter may help to provide further support for the notion that the suffering of individuals is the collective suffering of all.
References


