Title
Assessing Area-Level Suicide and Overdose Rates Via Google Search Term Data in the US

Author
Guerra, Christian

Publication Date
2022

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Assessing Area-Level Suicide and Overdose Rates Via Google Search Term Data in the US

Doctor of Philosophy

in

Sociology

by

Christian Guerra

December 2022

Dissertation Committee:
Dr. Bruce G. Link, Chairperson
Dr. David Brady
Dr. Wei Zhao

The Dissertation of Christian Guerra is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help, support, and assistance from the following people, institutions, and entities. I would like to thank the following: My mother and father, Erica & Hector. My advisor, Dr. Bruce G. Link. Committee members Dr. David Brady & Dr. Wei Zhao. Graduate advisor Dr. Tanya Nieri. A very special thank you to Ricardo Palacios, Dr. Bernie Berk, Tommy, Bella, Anna Wire, Sensei Kevin, Sensei Steven, the Aikido Club at UCR, the UCR Career Center, the UCR Student Recreation Center, the UCR Exploration Center for Innovative Teaching and Engagement (XCITE), Google, YouTube, The Getaway Cafe, CSU Los Angeles, and Los Angeles Valley College.

ABSTRACT OF THE DISSERTATION

Assessing Area-Level Suicide and Overdose Rates Via Google Search Term Data in the US

by

Christian Guerra

Doctor of Philosophy, Graduate Program in Sociology
University of California, Riverside, December 2022
Dr. Bruce G. Link, Chairperson

Suicide and fatal drug overdose are a major public health and social problem in the US. Google search term data may help researchers understand variations in area-level suicide and fatal drug overdose rates. The extent to which an area Googles about suicide and drug-related topics may be indicative of the overall social environment in the area. This project tests if there is an association between suicidal Google searches and the suicide rate, and also tests if there is an association between drug-related Google searches and the drug overdose rate. This project tests these associations using two-way fixed effects linear regression to control for unmeasured confounding, while controlling for the major predictors of each outcome. The unit of analysis are Designated Market Areas (n=209) and the analyses observe a seven-year period (2010-2016) for a total of 1,463 unique observations. Results show that suicidal Google searches are associated with suicide rates, and drug-related Google searches are associated with overdose rates. These findings demonstrate that Google searches may provide researchers with an additional lens into the innermost thoughts, attitudes, and behaviors of individuals at an area level. These findings suggest that Google search data can assist in real-time suicide and drug overdose monitoring and prevention.

TABLE OF CONTENTS

THEORETICAL FRAMEWORK	**1**
SOCIAL FIELDS	3
LINKING SOCIAL FIELDS AND GOOGLE SEARCHES	7
GOOGLE TRENDS	15
THIS PROJECT	19
POTENTIAL LIMITATIONS	24
CONTRIBUTION	26
ASSOCIATION BETWEEN SUICIDAL GOOGLE SEARCHES AND THE SUICIDE RATE	**29**
INTRODUCTION	29
HYPOTHESIS	46
DATA	46
METHODS	56
RESULTS	58
DISCUSSION	68
LIMITATIONS	73
POLICY IMPLICATIONS	75
ACADEMIC CONTRIBUTIONS	76
CONCLUSION	77
ASSOCIATION BETWEEN DRUG-RELATED GOOGLE SEARCHES AND THE DRUG OVERDOSE RATE	**78**
INTRODUCTION	78
HYPOTHESIS	93
DATA	93
METHODS	105
RESULTS	106
DISCUSSION	119
LIMITATIONS	123
POLICY IMPLICATIONS	124
ACADEMIC CONTRIBUTIONS	125
CONCLUSION	126

ASSOCIATION BETWEEN TRADITIONAL MASCULINITY AND FIREARM-RELATED GOOGLE SEARCHES AND THE MALE SUICIDE RATE **127**

INTRODUCTION 127
HYPOTHESES 145
DATA 146
METHODS 155
RESULTS 157
DISCUSSION 170
LIMITATIONS 174
POLICY IMPLICATIONS 175
ACADEMIC CONTRIBUTIONS 176
CONCLUSION 177

REFERENCES **178**

LIST OF FIGURES

THEORETICAL FRAMEWORK

Figure 1. Geographic distribution of Google searches containing the N-word 12
Figure 2. Geographic distribution of Google searches for “hunting” 13
Figure 3. Geographic distribution of Google searches for “rifle” 14
Figure 4. Geographic distribution of Google searches for “God” and “bible” 14
Figure 5. Frequency of Google searches for “porn” in a weekly time frame 17
Figure 6. Map of Designated Market Areas in the US 22

ASSOCIATION BETWEEN SUICIDAL GOOGLE SEARCHES AND THE SUICIDE RATE

Figure 1. Conceptual map of the study showing the direction of causality between variables 45
Figure 2. Map of Designated Market Areas in the US 47
Figure 3. Google search frequency for “suicide”, “kill myself”, and “I’m feeling suicidal” from June 27, 2021 to June 27, 2022 49
Figure 4. Correlation matrix of all covariates in the study 55
Figure 5. Prompt issued by Google when searching for suicidal topics 76

ASSOCIATION BETWEEN DRUG-RELATED GOOGLE SEARCHES AND THE DRUG OVERDOSE RATE

Figure 1. Number of annual deaths attributable to drug overdose in the US from 1999 to 2020 80
Figure 2. Conceptual map of the study showing the direction of causality between variables 93
Figure 3. Map of Designated Market Areas in the US 94
Figure 4. Google search frequency for “heroin”, “heroin addiction”, and “heroin addiction hotline” from June 27, 2021 to June 27, 2022 97
Figure 5. Correlation matrix of all covariates in the study 104
Figure 6. Graphic showing the fatal drug overdose rate through quartiles of the opioid search composite 114
Figure 7. Prompt issued by Google when searching for suicidal topics 125

ASSOCIATION BETWEEN TRADITIONAL MASCULINITY AND FIREARM-RELATED GOOGLE SEARCHES AND THE MALE SUICIDE RATE

Figure 1. Suicide rate in the US from 2000 to 2020 129
Figure 2. Differences in male and female suicide rates by age category 135
Figure 3. Geographic distribution of Google searches for “Red Sox” 143
Figure 4. Geographic distribution of Google searches for “bible” 143
Figure 5. Map of Designated Market Areas in the US 146
Figure 6. Correlation matrix of all covariates in the study 155

LIST OF TABLES

ASSOCIATION BETWEEN SUICIDAL GOOGLE SEARCHES AND THE SUICIDE RATE

Table 1. Bivariate linear regression on suicide rate (2010-2016) 59
Table 2. Multivariate linear regression on suicide rate (2010-2016) 61
Table 3. Descriptive statistics of covariates showing little change between 2010 and 2016 63
Table 4. Two-way fixed effects bivariate linear regression on suicide rate (2010-2016) 64
Table 5. Two-way fixed effects multivariate linear regression on suicide rate (2010-2016) 66

ASSOCIATION BETWEEN DRUG-RELATED GOOGLE SEARCHES AND THE DRUG OVERDOSE RATE

Table 1. Bivariate linear regression on fatal drug overdose rate (2010-2016) 107
Table 2. Multivariate linear regression on fatal drug overdose rate (2010-2016) 109
Table 3. Descriptive statistics of covariates showing changes between 2010 and 2016 110
Table 4. Two-way fixed effects bivariate linear regression on fatal drug overdose rate (2010-2016) 111
Table 5. Two-way fixed effects multivariate linear regression on fatal drug overdose rate (2010-2016) 113
Table 6. Lagged two-way fixed effects bivariate linear regression on fatal drug overdose rate (2010-2016) 115
Table 7. Lagged two-way fixed effects multivariate linear regression on fatal drug overdose rate (2010-2016) 117
Table 8. Lagged Arellano-Bond estimator with two-way fixed effects multivariate model Dependent variable: fatal drug overdose rate (2010-2016) 118

ASSOCIATION BETWEEN TRADITIONAL MASCULINITY AND FIREARM-RELATED GOOGLE SEARCHES AND THE MALE SUICIDE RATE

Table 1. Bivariate linear regression on male suicide rate (2010-2016) 158
Table 2. Multivariate linear regression on male suicide rate (2010-2016). Includes both Traditional Masculinity and Firearm search composites. 160
Table 3. Multivariate linear regression on male suicide rate (2010-2016). Includes only Traditional Masculinity search composite. 161
Table 4. Multivariate linear regression on male suicide rate (2010-2016). Includes only Firearm search composite. 162
Table 5. Two-way fixed effects bivariate linear regression on male suicide rate (2010-2016). 163
Table 6. Two-way fixed effects multivariate linear regression on male suicide rate (2010-2016). Includes both Traditional Masculinity and Firearm search composites. 165
Table 7. Two-way fixed effects multivariate linear regression on male suicide rate (2010-2016). Includes only Traditional Masculinity search composite. 166
Table 8. Two-way fixed effects multivariate linear regression on male suicide rate (2010-2016). Includes only Firearm search composite. 167

THEORETICAL FRAMEWORK

The United States has experienced substantial and sustained increases in rates of suicide and drug overdose that require a deeper understanding to guide policy and to inform interventions. Speaking to the deeply sociological nature of these trends, economists Case and Deaton (2015) have described them as one manifestation of “deaths of despair.” This being so, the proposed research identifies a potential lens, Google search behavior, as a means of probing the deeply private sentiments that are constitutive of this despair. Systematic research on Google search term data may assist social scientists and public health operatives to prevent suicides and fatal drug overdoses more efficiently than is now possible. The project rests on the fact that people are willing to share their innermost thoughts, attitudes, feelings, and inclinations when they conduct Google searches. Since Google is widely used, it has the potential to provide researchers with area-level cultures, attitudes, and behaviors. This project asserts that Google searches offer a potential lens on what has been called “the social field”, a phenomenon that is not directly observable that surrounds social actors and can only be identified through its effects. Because what people in an area Google about can give us a glimpse into the social field of an area, then we may be able to understand how these social fields drive individuals towards suicide or fatal drug overdose.

This study argues that social phenomena that are conducive to suicide, that form a part of an area’s social field, can be ascertained by examining what individuals in such area Google about. When an area undergoes certain stresses, anxieties, and problems, these issues form a part of the area’s social field, and are oftentimes reflected by high Google search volumes for topics related to these issues. For example, an area experiencing a high unemployment rate has a higher Google search volume for unemployment-related search terms (D'Amuri & Marcucci 2017). Following this logic, areas that undergo social phenomena that are conducive to suicide may Google more about suicide and suicide methods. Additionally, areas that undergo social phenomena that are conducive to drug use may Google more about drugs and ways to procure them. Therefore, **what an area tends to Google about frequently provides us with an indicator of the social field of such an area.**

This study will test whether social fields, reflected in Google searches, are associated with an area's suicide and fatal drug overdose rates, by examining if area changes in Google searches for suicide-related and drug-related topics are associated with changes in suicide rates and overdose mortality rates. This study will test if yearly changes in an area's Google search volume for suicide-related and drug-related content is associated with such an area's suicide and overdose mortality rate respectively. The years this study will observe are from 2010 to 2016. The yearly structure of the data also makes it possible to assess whether suicidal or drug-related Google searches are associated with the following year's suicide and overdose mortality rate. If such is the case, then it can be said that the social field of an area, reflected by Google searches, has some predictive power about the future behaviors of individuals in an area. The analyses will be conducted using fixed effects regression, which effectively controls for time-invariant factors such as the weather, elevation, culture, and sex. This study will also control for the popular suicide and fatal drug overdose predictors from the literature.

This chapter will first explain the theoretical framework of this project, social field theory. It will then explain how Google search data can assess area-level culture, attitudes, thoughts, and behaviors. Then it will cover the specifics of the project, its hypotheses, and its method of testing these hypotheses. Lastly, it will discuss the strengths and weaknesses of the project and its methods, along with the major contributions this study can potentially produce.

SOCIAL FIELDS

The theoretical framework for this project begins with Kurt Lewin's social field theory. Kurt Lewin was a prolific psychologist, widely regarded as the father of modern social psychology, who studied organizational change, group dynamics, and how individuals and their surroundings affect the behavior of others. In 1946 Lewin proposed the social field theory (Lewin 1948). Social field theory examines patterns of interaction and behavior between individuals and the surrounding environments where they move about (Burnes & Bargal 2017). Social field theory borrows its framework from the field of physics. Field theories in physics are used to explain how certain forces can have an influence on objects even if they are not in direct contact with one another. In the social world, forces in the field act on individuals and influence their thoughts, feelings, and behaviors in numerous ways even though there is no specific observable contact between individuals (Mohr. et al. 2020, 117). For example, a dramatic event in the social field, such as a terrorist attack in a train station, will change the way many individuals think about public safety, change the way individuals behave at train stations, and would change the way they travel, by canceling their train tickets. While one may not be directly affected by the terrorist attack, learning of the event and its specific details are more than enough to lead to changes in how an individual thinks and behaves in relation to train travel and public safety.

Every individual has a unique position in the social field possessing a unique magnitude of social force applied to them. This being so, while social fields can be said to be organized, the amount of force that is applied to individuals is not the same (Martin

2003). That is, individuals do not receive the same amount of force applied to them; there is a differential in the amount of force applied depending on the individual's position in the social field. The magnitude of social force applied to the individual is also not randomly distributed (Martin 2003). For example, a police shooting an unarmed black citizen undoubtedly causes disruptions and changes in the social field. However it has been found that black individuals report more poor mental health days after a racialized police shooting, while there is no change in poor mental health days for white individuals after the event (Curtis et al. 2021). This example illustrates how an event can apply a social force of anxiety and distress to a particular population, while leaving other populations relatively unscathed. Furthermore, this example illustrates how the application of social forces are not randomly distributed. In this case, it is more likely for minority populations to report poorer mental health after racialized police violence. Similarly, the pressure to commit suicide or engage in drug use is not randomly distributed in the US, there are areas with characteristics that make its inhabitants more likely to complete suicide or engage in drug use. We may be able to assess which areas predispose their inhabitants to succumbing to suicide or fatal overdose by observing the social field of these areas through Google searches.

Social fields provide individuals with a specific set of scripts, narratives, behaviors, and schemas (Mohr et al. 2020, 3). Social field theory asserts that behavior is derived from the totality of coexisting and interdependent forces that exerts an influence on an individual or group, and make up the social environment where the behavior takes place (Burnes & Cooke 2012). Because behavior is grounded in the totality of the present social environment an individual's behavior is a product of this totality, of all interactions and events around them. In mathematical terms, behavior (b) is a function of the interaction between a person (p) (or group) and their social environment (e) (Marrow 1977).

b = f (p, e)

Behavioral change is achieved by changes to the social forces that are present in the social field. For example, an individual (person) who recently learns that their company is enacting massive layoffs (environment), may be spurred to investigate what specific offices will receive the greatest number of layoff notices through a Google search (behavior). In another example, a foreign worker (person) who learns of a new work immigration law (environment) may investigate if this law applies to them via a Google search (behavior). Following this logic, an individual who is undergoing a divorce and suddenly loses their job, and exhausts their resources to ameliorate their dire situation, may be more likely to Google about escaping through suicide. In this way, negative social phenomena has influenced an individual to consider suicide and their considerations are recorded through their Google search activities.

Because social field theory pays attention to the way the totality of a social environment may be greater than the sum of its relational parts, social field theory focuses on larger forces that shape relationships within social networks. It also focuses on institutional forces that routinize social interactions and behaviors. Therefore, elements within a social field are recognized to be influenced by broader macro-level forces, regardless if researchers are able to measure these connections directly (Mohr et al. 2020, 116). For instance, **what people in an area Google about can be reflective**

of the total situation of such areas, informing us about the characteristics of such areas. For example, the areas that Google “God”, “bible”, and “Jesus” the most are heavily concentrated in the south. Using the sociological imagination we can therefore assess how much Christianity permeates the southern region’s politics, laws, institutions, attitudes, behaviors, and other aspects that make up its social environment. It turns out that the social environment of the south is heavily influenced by Christianity and in this way, we are able to understand how an area’s search behavior about religious topics can inform us not just about their religious beliefs, but on their overall social environment.

In a non-Google example, when US states passed laws banning same-sex marriage on a legal and institutional level, it was observed that LGB adults living in those states experienced a 37 percent increase in mood disorders, a 42 percent increase in alcohol use disorders, and a 248 percent increase in generalized anxiety disorders soon after the bans were implemented (Hatzenbuehler et al. 2009). LGB adults who resided in states without same-sex marriage bans did not experience any significant increases in psychiatric disorders during the same period. Further supporting evidence regarding the effect of these bans shows that the mental health of heterosexual adults in states that had bans remained largely unchanged during the study period. This finding supports the notion that larger institutional forces have an effect on the social field of an area, and that these effects are not randomly distributed, but are applied differentially throughout the population. Just as there is no known field in physics to affect all particles, some individuals may not be affected by these changes (Martin 2003), as in the example of heterosexuals in states before and after gay marriage bans.

While critics of social field theory note that social fields cannot be seen directly, such is the case with several phenomena such as gravity, electric fields, and magnetic fields. All such phenomena can be understood with specific tools that assist in uncovering knowledge about the given phenomena. While social fields cannot be directly observed, they can be measured by observing the effects that changes in the social field have on populations. For example, noted in the previous paragraphs, the change in the social field would be passing legislation banning same-sex marriage. The effects are the increasing rates of mood disorders, substance use disorders, and anxiety disorders among LGB residents.

There is then good reason to believe that a social field can be measured via what people search for on Google. Essentially, **the effects that individuals experience from changes in the social field show up in Google searches**.

LINKING SOCIAL FIELDS AND GOOGLE SEARCHES

When a change occurs in the social field, the thoughts and behaviors of individuals are affected. Often, these thoughts and behaviors can show up as Google searches. A change or disruption to the social field, such as a snow storm, may influence many individuals to Google about the "highway closures" near them. Therefore, through Google searches, we can observe how individuals are affected by changes in the social field. In another example, a celebrity suicide produces a change in the social field. News of a celebrity spreads quickly through news and social media outlets, which then affects individual thoughts and behaviors. After a celebrity commits suicide, the effects are observable spikes in the frequency of Google searches related to suicide. Consistent with this reasoning , several studies report a higher incidence of suicides soon after a celebrity suicide (Jeong et al. 2011, Ju et al. 2014). Therefore, while the social field cannot be seen nor measured directly, it is observable through its effects (Martin 2003). The change in the social field in this case is a celebrity suicide, and the effects are increased levels of Googling for suicide-related topics. Therefore, the effects of changes in the social field can be measured by observing and tracking Google searches.

Tracking the frequency and location of Google searches has produced promising results. The first study using Google tracked outbreaks of the flu much more effectively than the CDC could (Ginsberg et al. 2009). A change in the social field, a flu outbreak, influences the behavior of residents in an area, and the resulting thoughts, concerns, and behaviors can be observed by what these areas Google about. It was found that when an area experienced a flu outbreak, residents in these areas would Google for “flu remedy”, “flu symptoms”, and several other flu-related terms. In this instance, the flu outbreak is a change in the social field that can be measured and geolocated, thanks to the Google search behavior of the area’s residents.

More recently, COVID-19 has created enormous changes to the social field with lasting effects on health, surveillance, travel, and labor structures. Several of these changes are reflected in what people search on Google, where they search it from, and how frequently terms are searched. For example, similar to Ginsberg et al. 's 2009 flu study, Mavragani & Gkillas (2020) found that areas experiencing COVID-19 outbreaks were Googling for COVID-related topics. Indeed, when changes to the social field take place, individuals react by informing themselves on Google, the world's largest source of information.

In addition to COVID-related searches being able to track COVID-19 outbreaks, Gimbrone et al. (2021) found that during the first pandemic wave in early spring 2020, Google searches for economic stressors skyrocketed. The study also found that during the pandemic there was a significant increase in the frequency of mental health-related and suicide-related Google searches in the US. These findings are not surprising as the prolonged social isolation and economic distress from the pandemic lockdowns took a heavy toll on the mental health of Americans. In this instance, the major change to the social field, COVID-19 and its lockdowns, spurred millions of Americans to Google about the virus, their mental health symptoms, their economic woes, and ways for them to escape their unfortunate situations via suicide.

Ginsberg et al. 's 2009 landmark study helped establish the credibility of Google search data in health research. Since then, there has been a 20-fold increase in studies using Google search data (Arora et al. 2019). These studies have been able to successfully track, predict, or monitor the following health topics: herpes, herpes zoster vaccinations, outbreaks of the plague, cancer, whooping cough, influenza, asthma, Lyme disease, tropical infections, syphilis, HIV, influenza hospitalizations, systematic lupus erythematosus, depression, obesity, West Nile virus, Zika virus, AIDS prevalence, suicide, drug use, and dangerous pharmaceutical drug interactions (Mavragani & Ochoa, 2019; Arora et al. 2019, Jun et al. 2018, Brodeur et al. 2021)

In addition to tracking disease and its effects, Google search data reveals social patterns that arise from events. In other words, **changes in the social field can be captured by what people Google after the change.** For example, searches for “will I be deported” rose precipitously after the election of Donald Trump (Chykina & Crabtree 2018). Searches looking for off-the-books ways to terminate a pregnancy are higher in states that have passed laws restricting abortions (Stephens-Davidowitz 2017). Searches for terms such as “my mom beat me” and “my dad hit me” increased during the Great Recession and were closely tracked with the unemployment rate (Stephens-Davidowitz 2017). Given that an overwhelming amount of child abuse cases go unreported, and social service agencies were understaffed because of the recession, reports of child abuse failed to observe this disturbing trend. Google searches, on the other hand, were able to shed light on this difficult-to-observe phenomenon. In similar ways, an individual struggling with suicidal thoughts or with a drug addiction may search for topics relating to their stressful situation. Therefore, **Google search data has the capacity to reveal to us the unfortunate and painful realities many individuals find themselves in.** The startling findings above demonstrate how Google searches reveal meaningful social patterns and phenomena across geographic areas.

In addition to social patterns, Google search data can reveal meaningful cultural norms and social attitudes that are reflective of the social field of areas. Culture is a large determinant of an individual's attitudes, thoughts, and behaviors. **Because area-level culture and attitudes influence residents to Google about certain topics more frequently compared to other areas, we are able to observe the underlying social field of an area by examining the Google search behaviors of such areas.** Capturing the geographical pattern of those subscribing to certain cultures, attitudes, or behaviors may be useful to social scientists. It is important to be able to assess area-level culture, attitudes, thoughts, and behaviors because several of these factors have far-reaching consequences for those who subscribe to these factors.

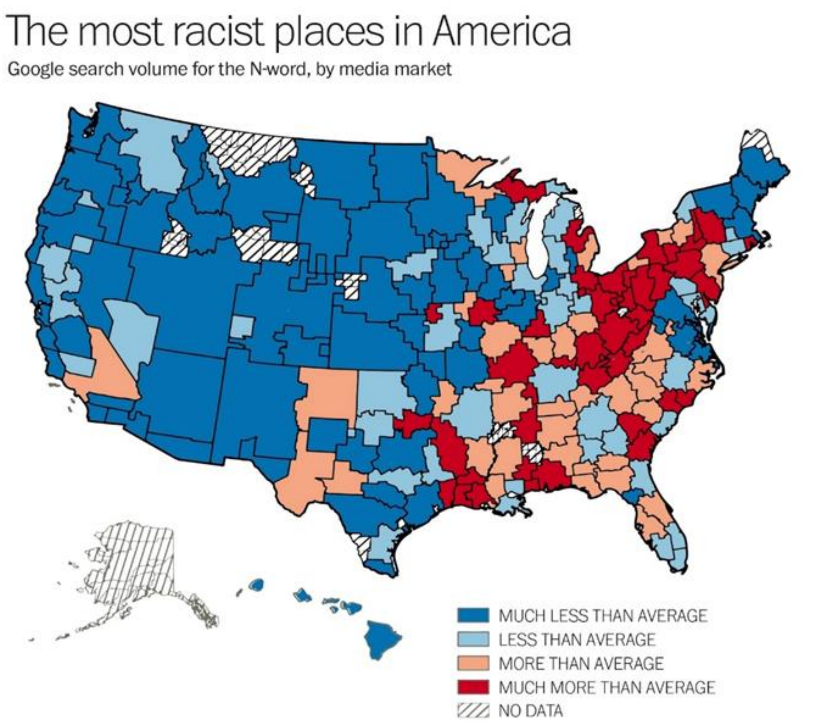
Regarding social attitudes, recently there has been an interest in assessing area-level racism and prejudice using Google search data. Racist and prejudicial attitudes are difficult to assess using traditional survey and interview methods. Survey data suffers from the possibility of social desirability bias where the respondent portrays themselves in a more positive light than true responses might reveal. Survey data also suffers from interviewer bias, where the interviewer judges a respondent based on their characteristics, which then may elicit erroneous or biased answers. Survey data suffers from the observer-expectancy effect, where a researcher's biases cause them to influence the behaviors or responses of a subject. Generating a reliable and representative measure of area-level attitudes through surveys can be costly and labor intensive.

However, because the internet provides some sense of anonymity, it increases the likelihood of expressing prejudicial and discriminatory attitudes (Fox & Rainie 2014, McKenna & Bargh 2000). Several studies have argued that Google search data could elucidate racist and prejudicial attitudes more clearly (Chae et al. 2015, Chae et al. 2018). For example, we can expect an intolerant and prejudicial area to search for more racially-charged terms. Through a social field perspective, because an individual knows that their surrounding area is not supportive of them, it influences their thoughts and behaviors, often leading to negative outcomes. Therefore, we can expect minorities to experience poorer health in intolerant and prejudicial areas, or areas with a negative social climate.

Some studies have observed the effects of area-level racism and prejudice by using Google search data. Chae et al. (2015) found that market areas that frequently Googled the racist n-word were associated with an 8.2 percent increase in the all-cause

Black mortality rate in that area. Using the same independent variable, Chae et al. (2018) found that each standard deviation increase in area racism was associated with a 5 percent increase in the prevalence of preterm births among Blacks, and also associated with a 5 percent increase in the prevalence of low birth weight among Blacks. An illustration showing which areas search the n-word the most is shown below (Chae et al. 2015).

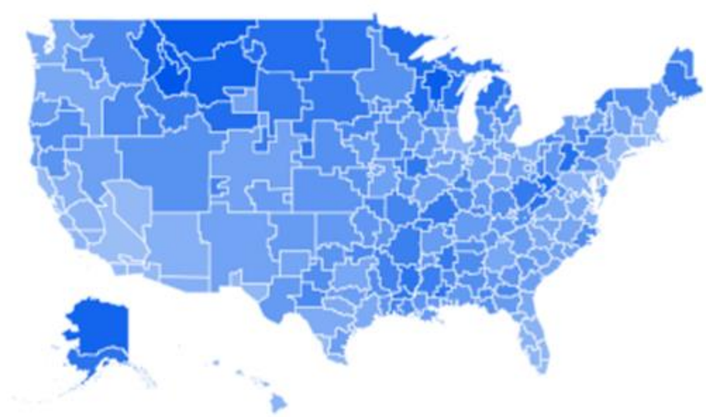
Figure 1. Geographic distribution of Google searches containing the N-word (2004-2007)



Lastly, the higher the racially-charged Google search rate in an area, the worse Barack Obama performed in the 2008 election in that particular area (Stephens-Davidowitz 2014). Therefore, geographic area social attitudes can be assessed using Google search data, and research has shown its effects on those residing in these areas.

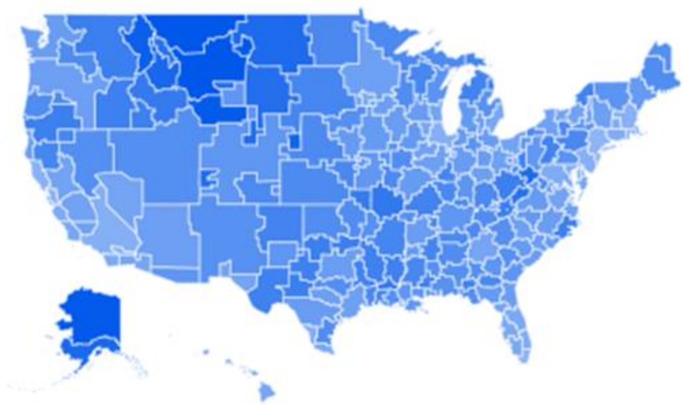
In similar ways, **activities, behaviors, and interests of an area can indicate the social attitudes and cultural norms of an area, which are indicative of the social field.** That is, the popularity of certain interests and topics of an area are indicative of characteristics of that area's population. For example, if we assess the popularity of the search term "hunting" by geographic area (shown below), we observe that the general interest for hunting is highly concentrated in the central-north, Great Lakes, and Alaska regions.

Figure 2. Geographic distribution of Google searches for "hunting" (last 5 years)



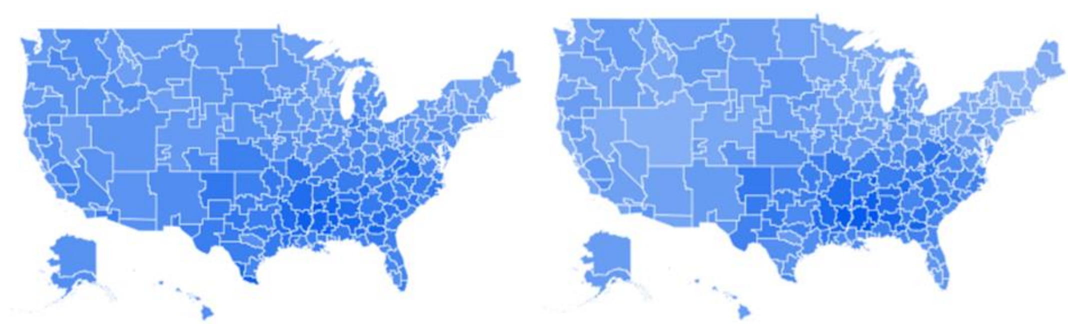
From this fact, we can infer that these areas are very likely to show interest in hunting gear such as rifles and outdoor equipment. The popularity of Google searches for "rifle" (shown below) show us this is the case, as both maps for "hunting" and "rifle" have much overlap.

Figure 3. Geographic distribution of Google searches for "rifle" (last 5 years)



In addition to interests and activities, are we able to assess cultural phenomena with Google Trends? Mapping out the popularity of the search term "God" (figure 4 left) and "bible" (figure 4 right) shows us this may very well be the case. A high concentration of interest in these terms correlates with the fact that the US south has high rates of religiosity.

Figure 4. Geographic distribution of Google searches for "God" (left) and "bible" (right) (last 5 years)



Therefore, an area's interest in activities and topics may in fact be indicative of underlying characteristics of that area's population. Google search data allows us to observe the popularity of activities, behaviors, and interests by geographic location.

Because some of these activities and behaviors may be associated with a higher rates of suicide, such as firearm ownership (Knopov et al. 2018, Kposowa et al. 2016, Vitt et al. 2018), it may be useful to test whether firearm-related search terms are associated with suicide rates. Additionally, if there is a culture that is conducive to male suicide, such as traditional masculinity (Coleman et al. 2011, Houle et al. 2008), it may be useful to test whether traditional masculinity-related search terms are associated with the male suicide rate. In this way, it is possible to test if areas with high search interest for firearms, for example, have higher rates of suicide.

In fact, this project will do just that, in addition to examining if suicide-related Google searches are associated with suicide rates, and if drug-related Google searches are associated with fatal drug overdose rates, this project will examine if firearm-related and traditional masculinity- related Google searches are associated with the male suicide rate. There is an epidemiological framework that links firearms to male suicide, and there is a sociological theoretical framework that links traditional masculinity to male suicide. These frameworks are discussed in Chapter 3 of this project.

Because activities, behaviors, and interests oftentimes emanate from the social field, and we are able to assess the geographic interest of these activities, behaviors, and interests through Google searches, this search data may provide us with a glimpse of the social field of these areas.

GOOGLE TRENDS

Google search data can be obtained via Google Trends. Google Trends is a data tool that provides real-time and historical data regarding Google searches dating as far back as 2004. Google Trends can provide the frequency that a term has been searched thereby gauging the popularity of such a term. Additionally, Google Trends provides data on the location of these searches by country, state, city, and the Designated Market Areas where these searches take place. Therefore, we can observe the geographic areas where a search term is more popular. This project will focus on Designated Market Areas, the smallest geographic tract available on Google Trends. Additional information regarding Designated Market Areas is found later in this introductory chapter.

What makes Google Trends a novel data source is the fact that the data are collected and reported in real-time, something public health datasets are unable to do (Arora et al. 2019). Since the data are provided in real-time, it is a tremendous improvement over more traditional health surveys such as the CDC's Behavioral Risk Factor Surveillance System (BRFSS) which takes over a year to assess and analyze. Therefore, **we may be able to assess which areas have more citizens contemplating suicide in real-time by surveying suicide-related search terms. Likewise, we may be better able to assess which areas are more prone to fatal drug use in real-time by surveying drug-related search terms.** These real-time strategies would give public health officials a timely opportunity to implement suicide or overdose prevention programs, instead of having to wait months for the latest mortality data.

Of course one can imagine some reluctance about using Google search data. How can seemingly random Google searches be turned into useful behavioral insights? From the examples listed above it is evident that Google searches are not random; they are patterned geographically and through time. For example, patterns in Google search behavior can be observed through time by looking at the frequency of the search term "porn" over the last seven days in the US (below). The seven peaks (one for each day) shown on the graph below demonstrate that 1 AM is the time when most people are searching for porn on Google. These patterns are remarkably consistent, uniform, and can be replicated at any date. This demonstrates that social behavior is reflected in what they Google about and when they Google about it.

Figure 5. Frequency of Google searches for "porn" in a weekly time frame (July 3-9, 2022). Frequency peaks daily at about 1am.

There is also some critique that Google is simply not used enough to produce meaningful insights. However, overwhelmingly, the most popular internet resource in the world is Google (Jun et al. 2018). What also makes Google a very attractive data source is the fact that Google has over 90 percent of the market share in online searches in the US. Google is the world's most utilized search engine, handling roughly 2 trillion searches per year, 167 billion searches per month, and 5.5 billion searches per day (Jun et al. 2018), making it the world's largest data source available. Eighty percent of internet users in the US have searched for health-related topics on Google (Pew 2013). Given these usage statistics, Google search data is simply the best online search behavior data we currently have.

The power of Google search data lies not only in the frequency of its use, but in its ability to capture what are often hidden thoughts, attitudes, inclinations, and behaviors of its users. Google search data can reveal profound truths about the human condition at any particular place and time. The power in these data lies in the fact that individuals often search for topics they dare not ask their fellow peers, neighbors, and spouses about. An individual eager to learn about their sexless marriage, mental health issues, deep insecurities, or prejudices are free to do so online, without facing any social repercussions given the anonymity the internet confers (Stephens-Davidowitz 2017). Not many individuals would admit to a researcher that they regret having their children, yet this sentiment is frequently Googled, leading to numerous forums discussing this taboo subject (Stephens-Davidowitz 2017, Moore & Abetz 2019).

It turns out that individuals are more than willing to query about sensitive topics that are too embarrassing or uncomfortable to ask and receive advice and information about from others. Controversial topics that would otherwise result in one being fired, ostracized, or stigmatized are free to be searched, without fear of societal repercussion or backlash. In many ways, Google search data provides a “digital truth serum” yielding information that would otherwise never be disclosed to anybody (Stephens-Davidowitz 2017). In relation to this project, an individual who is experiencing suicidal thoughts may research this topic via Google, before they may want to confer with their family and peers. In the same respect, an individual with a substance use disorder may seek help online before they notify their peers. **Because Google search data captures instances of an individual’s inner-most thoughts, sentiments, and attitudes that undoubtedly influence their behaviors, Google search data may provide social scientists with a glimpse into the effects of the social field. That is, changes to the social field, such as the construction of a toxic factory next to a neighborhood, may lead the surrounding neighborhood residents to Google about their emerging lung condition.**

THIS PROJECT

Spring boarding from these perspectives and findings, this project investigates:

1 Whether there is an association between suicide-related Google search terms and area-level suicide rates

2 Whether there is an association between drug-related Google search terms and area-level fatal drug overdose rates

3 Whether there is an association between firearm-related Google search terms and area-level male suicide rates

4 Whether there is an association between traditional masculinity-related Google search terms and area-level male suicide rates

In addition to testing these associations on a yearly basis, this project will also test if a previous year's Google searches has an effect on the next year's mortality rates. This would test if Google searches have some kind of predictive power.

Just as previous projects have found that racist Google searches can provide us with an insight into the social field of such areas, this project proposes to test the possibility that Google searches can provide us with an insight into the social field of an area, and whether this social field is associated with the rates of suicide and fatal drug overdose. More specifically, areas that frequently search for content that informs them on how to complete a successful suicide should have higher rates of suicide. Areas that frequently search for drug-related content should have higher rates of drug overdose deaths. Areas that frequently search for firearm and traditional masculinity-related content should have higher rates of male suicide.

The dependent variables in this study, suicide rates and drug overdose mortality rates, are obtained from the CDC's Compressed Mortality File. The CDC's Compressed

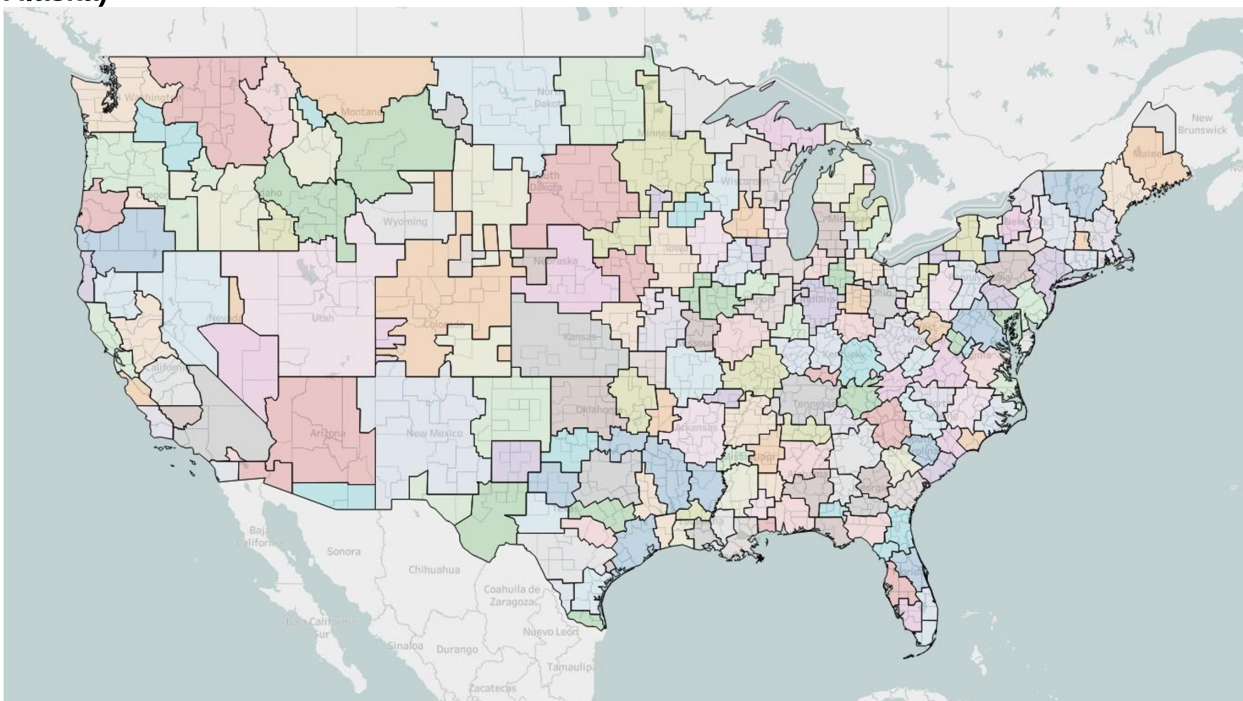
Mortality File is an appropriate source of data for this project given its reliability to provide detailed mortality data. This file provides data on every death in the US in a yearly format. Every entry in the dataset provides the cause of death, age, sex, race, and county of each individual. This study will use 7 years of data spanning from 2010 to 2016, the last year of data provided by the dataset. Since the data is provided at the county level, county data are aggregated to the Designated Market Area-level using county population as weights. Since there are 210 Designated Market Areas in the US, and 7 years of data, the number of observations of this project is approximately 1470. Suicides and drug overdose deaths will be subset from the data to create rates. The variable of interest, Google searches, will be linked to suicide and drug overdose rates, an objective outcome from a reliable data source.

As the project aims described above will be conducted while keeping in mind previous research that has already tested these associations. Previous research on the association between suicide-related Google searches and suicide rate has been successfully observed in Japan, South Korea, Taiwan, Germany, Spain, and several other countries (Hagihara et al. 2011, Sueki 2012, Song et al. 2014, Chang et al. 2015, Paul et al. 2017). However, very few studies have examined this association in the US, which is surprising given the precipitous rise in suicides over the last two decades.

Using Pearson correlations, Gunn & Lester (2013) found that Google searches for “commit suicide”, ”how to suicide”, and “suicide prevention” were positively associated with the 50-state suicide rate in the year 2009. Using a non-linear lasso model, Parker et al. (2017) found that life stressor-related Google searches were able to better predict the 50-state suicide rate than alternative models for the year 2015. Using ARIMA models, Capron et al. (2021) found that firearm suicide-related Google searches are associated with the US national-level suicide rate from 2004 through 2016. Using Box-Jenkins transfer function models, Lee (2020) found that anxiety, sleep, and unemployment-related terms showed a significant correlation to the national-level suicide rate for 2004 through 2017. Using mixed model analysis, Adam-Troian & Arciszewski (2020) found that absolutist words such as “completely” and “totally” are predictive of the 50-state suicide rate from 2004 through 2017. Using fixed effects regression, Guerra (2019) found that a Google search term composite scale comprised of search terms “suicide”, “suicide prevention”, “cyanide”, “suicide hotline”, and “suicidal” are significantly associated with the 50-state suicide rate from 2006 to 2016.

Given the review above, numerous studies have observed an association between suicide-related Google searches and the suicide rate on a national and state-level. However, what has yet to be done is testing this association on a more granular scale. This project will do so at the smallest geographical tract available by Google Trends, the Designated Market Area (DMA). This is the first study to test this association at the DMA level. There are 210 DMAs in the US. DMAs are designed for media marketing purposes and thus individuals within these DMAs are expected to share similar characteristics, interests, and profiles. An illustration of the DMAs is shown below.

Figure 6. Map of Designated Market Areas in the US (Not pictured: Hawaii & Alaska)



In addition to suicides, this project will also focus on fatal drug overdoses. Thus far there have been very few studies that have attempted to test if there is an association between drug-related Google searches and fatal drug overdoses. Using an Extremely Random Forest Machine Learning model, Campo et al. (2020) found that the following Google terms were associated with the 50-state fatal drug overdose rate from 2004 to 2017: Narcan, heroin, opioid, Vivitrol, suboxone, naloxone, perc, overdose, methadone, and withdrawal. Using linear regression, Mukherjee et al. (2020) found that Google searches for drug names such as “heroin”, “fentanyl”, “oxycodone”, and “cocaine” were useful in forecasting weekly drug overdoses in the state of Connecticut from 2012 to 2018. Using a hierarchical regression model, Arendt (2021) found that the Google search term “fentanyl” was associated with the 50-state fatal drug overdose rate from 2004 to 2017. These are all the studies done so far on this topic.

Given the dearth of knowledge on this subject, this study will explore if there is an association between drug-related Google searches and the fatal drug overdose rate at the DMA-level. While there have been only a few studies investigating this association, this project goes further by doing so at the DMA-level, a level not yet examined. This is the first study to analyze this association at the DMA level.

Lastly, if area-level Google searches for racially-charged terms are associated with negative political and health outcomes for African-Americans, then it is indicative that area-level Google searches provide a window into the social field of such areas. This project argues that the culture of Traditional Masculinity, which is conducive to male suicide (Kalish & Kimmel 2010), may be observed through Google searches through Traditional Masculinity-related topics, behaviors, and attitudes. There has yet to be a study that assesses a culture through Google Trends. Similar studies have assessed area-level intolerance and racism using the terms “kkk”, “hitler”, and the n-word (Chan 2019, Chae 2015), however this project aims to gain a broader and deeper understanding of the culture of Traditional Masculinity by using a far greater number of terms.

Lastly, this project will test if firearm-related Google searches are associated with the male suicide rate. Thus far there have been two studies testing this association. Using ARIMA models, Capron et al. (2021) found that the terms “how to buy a gun”, “where to buy a gun”, “where to buy guns online”, “where to buy used guns”, and “deadliest bullet” showed consistent associations with the national suicide rate on a monthly time structure. Using transfer function models, Lee et al. (2021) found that the term “gun suicide” was significantly associated with the national suicide rate on a monthly time structure.

In summary, this project will consist of three chapters. They are the following:

Chapter 1 - The association between suicide-related Google searches and the suicide rate (2010-2016)

Chapter 2 - The association between drug-related Google searches and the fatal drug overdose rate (2010-2016)

Chapter 3 - The association between Traditional Masculinity & Gun-related searches and the male suicide rate (2010-2016)

POTENTIAL LIMITATIONS

While there is widespread internet usage among Americans, sociologists understand that technologies and amenities are not distributed evenly among populations. There are documented disparities in internet usage among age groups, educational attainment, income, race/ethnicity, and urbanity (Statista 2021). While internet usage ranges from 99 percent to 96 percent for the ages of 18 to 64, only 75 percent of adults over the age of 65 use the internet.

Generally, there are no observed gender disparities in internet usage or access. The share of college graduates and those with some college who use the internet ranges between 98 to 97 percent however, the share of those with high school or less who use the internet is 86 percent. Given the correlation between education and income, it is no surprise that 86 percent of those who report less than 30,000 dollars yearly income are internet users, compared to 98 percent of those who report making over 50,000 dollars a year (Statista 2021).

In terms of race/ethnicity, surprisingly Hispanics use the internet slightly more than whites at 95 and 93 percent respectively. 91 percent of African Americans are internet users. An explanation for high Hispanic internet usage most likely lies in the increase of the young Hispanic population and the rapid aging of the white population in the US. In terms of geographic location and environment, 95 percent of urban residents are internet users, compared to 94 percent of suburban residents, and 90 percent of rural residents.

Given these disparities, researchers should acknowledge that these methods may not provide complete data, especially on older populations, the less educated, those of low socioeconomic status, some ethnic and racial minorities, and those residing in rural areas. However, given these disparities in internet usage, the internet is becoming increasingly affordable and available as time goes on. The share of Americans who did not use the internet has declined drastically from 48 percent in the year 2000 to just 7 percent as of 2021. As more Americans actively use the internet throughout the years, there is an increasing incentive to utilize internet data for all sorts of research purposes.

Google Trends should not be expected to supplant traditional interview and survey methods, but should be seen as another tool to uncover hidden sentiments in society. Google Trends is a useful tool to assess an area's attitudes and inclinations, however it cannot be used at the individual level. Assessing an individual's predisposition towards suicide, for example, is best done through traditional survey or interview methods. However, assessing a geographical area's predisposition towards suicide can be best measured using suicidal-related Google search terms. Similarly, an individual's experiences with racism or discrimination are best assessed through surveys or interviews. However, assessing the levels of racism and intolerance in a geographical area can be better measured using racist and hateful Google search terms. For these reasons, Google Trends is especially useful in measuring the social field within a geographical area.

Lastly, not all Google searches for suicide-related content are necessarily made by people contemplating suicide or looking for suicide methods. Nonetheless, an area conducting a more than average search volume for suicide-related content can be said to be reflecting the overall social field of such area. For example, Googling for “COVID symptoms” does not necessarily mean one has COVID, but it does reflect a sentiment that is influenced by the surrounding social field. Changes in the social field affects social environments, albeit not equally, however the distribution of these effects can be assessed via Google searches. Therefore, if an area experiences a COVID outbreak, of course not all COVID-related Google searches from this area are done by those who have contracted COVID, but rather these searches reflect a common sentiment or anxiety that is shared by the surrounding social environment.

CONTRIBUTION

This project makes several contributions to several fields and to public health. Firstly, this project advances our theoretical and methodological understanding of social fields. This project argues that social fields may be measured through what people search on Google. This is the first time any project argues or tests this idea.

This project advances Google Trends research by implementing the analysis at the most refined geographic tract available. Most studies using Google Trends are done at the national and state-level, however this is one of the few studies that is done at the DMA-level. Furthermore, this project will attempt to use several search terms and analyze them via principal component analysis to construct reliable scales, whereas previous studies only use a few individual search terms. This is an improvement from previous studies using this method because those studies have analyzed DMA-level associations with only a few terms. For example, Chan (2019) used the search term “Hitler” and ‘kkk” to assess DMA-level intolerance, Hall et al. (2020) used the search term “pollen” to assess seasonal patterns in DMA-area level pollen concentration, Cousins et al. (2020) used a few COVID-19 -related search terms such as “am I sick”, “covid symptoms”, and “covid testing” to assess DMA-level COVID-19 cases, and Chae et al. (2018) used the n-word search term to assess DMA-level racist social climates. This project argues that a more comprehensive and exhaustive list of search terms may be better in assessing social phenomena such as suicide and drug overdose.

This project advances epidemiology and suicidology research by implementing an innovative methodology that is seldom used. Previous studies have validated this methodology at the national and state-level for suicide and fatal drug overdose outcomes. However, no previous studies have done these analyses at the DMA level for either health outcome. This project intends to do so. Furthermore, current suicidology research heavily focuses on individual-level risk factors and overlooks larger structures such as cultural and social climatic factors, which this project analyses.

This project advances sociological research in that it introduces an emerging and verified methodology to the field. Sociologists have not identified innovative and sociologically interesting approaches to suicide that are available in other disciplines (Wray et al. 2011). Current sociological research can largely benefit from up-to-date methods that can complement existing methods such as surveys and interviews. This project shows the importance of not focusing on what people say, but rather on what they think. Google can show us people’s innermost thoughts, and this project shows sociologists that this data ought to be taken advantage of. Furthermore, this project revives a long sociological tradition of studying suicide. Since 1980, of the 30,000 academic articles on suicide, only 1.3 percent were categorized as sociological (Wray et al. 2011). Given the dearth of recent sociological research on suicide, this project fills many gaps in the methods of studying suicide.

Lastly, this project can help inform public health initiatives aimed at reducing suicide and assessing the prevalence of drug use. The geographical nature of this project may further inform public health officials and policymakers on where to target drug overdose and suicide prevention interventions. Areas that frequently Google drug-related and suicide-related content can be assessed in real-time thus, improving current methods that rely on previous data.

Association between Suicidal Google Searches and the Suicide Rate

Abstract: Suicides have been increasing dramatically in the last two decades. Google search term data may help researchers understand variations in area-level suicide rates. This analysis tests whether suicide-related Google searches are associated with area-level suicide rates from 2010 through 2016 using two-way fixed effects as a means of controlling unmeasured confounding. The analysis controls for major suicide predictors from the sociological, epidemiological, psychiatric, and suicidology literature and includes area-level marriage, divorce, veteran status, living alone, not having health insurance, poor health, heavy alcohol use, physicians per capita, and mental health providers per capita. Results demonstrate some suicide-related search terms are associated with area-level suicide rates.

INTRODUCTION

Suicide is increasingly becoming a public health issue in the US. While several predictors from the literature provide reliable assessments of area-level mortality, these assessments are useful after suicides have occurred. This study proposes using a new lens of observing social phenomena, real-time Google search term data, to understand variations in area-level suicide rates. Google provides information on how popular search terms are in areas over time. Therefore, **it is possible to test if area-level suicidal Google queries are associated with the suicide rate**. And, if this association exists, then it opens the possibility to assess suicide-related Google search behavior in real time.

Furthermore, this project will test if Google searches allow us to gain further insight into the effects of the social field. The social field is an integral part of social field theory, which examines patterns of interaction and behavior between individuals and the surrounding environments where they move about (Burnes & Bargal 2017). Social field theory is used to explain how social forces can have an influence on individuals even if they are not in direct contact with one another. Social forces in the social field act on individuals and influence their thoughts, feelings, and behaviors even though there is no specific observable contact between individuals (Mohr. et al. 2020, 117). The social field can be measured through the effects produced by changes in the field. For example, when US states ban abortion, which is a change in the social field, it has been observed that these states Google more frequently for off-the-book ways of obtaining an abortion (Stephens-Davidowitz 2017).

In another example, Google searches for “will I be deported” rose precipitously after the election of Donald Trump (Chykina & Crabtree 2018). According to social field theory, the field can only be measured through the effect of changes in the social field. Therefore, the effect of Donald Trump being elected can be observed by the rise in searches for “will I be deported.” These examples demonstrate that Google search data has the capacity to reveal to us the unfortunate and painful realities many individuals find themselves in. As mentioned above, this project tests if area-level suicide-related Google searches are associated with suicide rates using longitudinal data spanning seven years.

This paper begins with a review of the suicide problem in the US and the major predictors from the sociological, epidemiological, psychiatric, and suicidology literature. It then describes the utility of using Google data and why it is a good and novel source of data. This section is followed by a review of previous studies that have tested the association between suicide-related Google searches and suicides. That section is followed by the major contributions of this project, followed by the data, methods, results, and discussion sections.

Suicide is a Major Public Health Problem

The suicide rate in the US has increased by 35 percent since the year 2000 (CDC 2020). Suicide is the second leading cause of death for individuals between ages 10-14 and 25-34 and the third for ages 25-34 (NIMH 2022). In 2018, almost 11 million Americans seriously contemplated suicide, 3.3 million made a plan, 1.4 million attempted suicide, and nearly 50,000 completed suicide (CDC 2020). A life is lost to suicide every eleven minutes in the US (CDC 2020). Individual-level pathways to suicide include divorce (Neumayer 2003, Wray et al. 2011), unemployment (Kposowa et al. 2019, Neumayer 2003, Norstrom 1995), depression (Phillips et al. 2007), stress (Lester & Gunn 2016), and alcohol abuse (Neumayer 2003, Norstrom 1995, Kaplan et al. 2016). However, these predictors have produced mixed results in recent years (Wray et al. 2011). Having a history of previous attempts is one the most prominent and consistent predictors of suicide, however this data is difficult to procure (Owens et al. 2002).

More consistent area-level pathways to suicide include firearm ownership, having a firearm in the household, and storing firearms unlocked (Knopov et al. 2018, Kposowa et al. 2016, Vitt et al. 2018). Given that over half of suicides in the US are conducted via firearm, firearm ownership data provides a consistent estimate of suicides and firearm suicides in the US. Additional risk factors include psychiatric, substance dependency, and substance use disorders (Miller et al. 2012)

Excessive alcohol consumption is associated with suicide (Kaplan et al. 2016, Neumayer 2003). A meta-analysis found that acute alcohol use (consuming excessive amounts prior to a behavior) is a major risk factor for attempting suicide (Borges et al.

2017). Kaplan et al. (2013) found that almost 25 percent of men and 17 percent of women who had completed suicide were found to have alcohol blood concentration. Alcohol intoxication may facilitate completed suicide by disinhibiting or diminishing an individual’s aversion and fearful response to self-harm. Additionally, chronic and heavy alcohol consumption may be an indicator of underlying mental illness, which is tied to suicide (Miller et al. 2012).

Chronic pain, regardless of type, is a risk factor for suicidality (Racine 2018, Hassett et al. 2014). Physical health conditions are found to be associated with higher odds for reporting suicide ideation, a suicide plan, and suicide attempts (Stickley et al. 2020). It is hypothesized that chronic pain may facilitate the development of depression, hopelessness, increase the desire of escape by death, and erode the fear of dying (Hooley et al. 2014). Chronic pain may create severely negative physiological and psychological conditions that an individual may be incentivized to escape from via suicide. Chronic pain may also increase an individual's tolerance of pain, making it easier to endure painful processes while attempting suicide.

Being of veteran status is a risk factor for suicide (Cerel et al. 2015, McCarten et al. 2015). It is hypothesized that military service may decrease an individual’s perceptions of danger, self-harm, and fear of death, thereby increasing one’s chances of attempting suicide (Joiner 2005). Additionally, combat exposure is hypothesized to contribute to fearlessness and habituation to death (Wolfe-Clark & Bryan 2017). A recent prominent empirical finding is the prevalence of physical health conditions among veterans who have committed suicide (Wood et al. 2020). This study indicates that military service has a high rate of physical injuries while military culture values physical strength and capacity. Being injured while participating in a culture that values physical strength may produce a sense of “perceived burdensomeness” that may be conducive to suicide (Joiner et al. 2009). Additionally, veterans experience high rates of posttraumatic stress disorder and major depressive disorder, which is tied suicide (Nichter et al. 2019)

The classical sociological risk factors for suicide which includes divorce, unemployment, and marriage produce mixed associations with suicide (Norstrom 1995). Using the National Longitudinal Mortality Survey, Kposowa (2000) found that from 1979 to 1989, higher rates of suicide were found in the divorced. Using the same dataset, Kposowa et al. (2020) found that from 1990 to 2011, the divorced and separated are 88 percent more likely to complete suicide than their married counterparts. Recently, unemployment was found to be significantly associated with suicide (Kposowa et al. 2019). However, these factors produce mixed results as consistent findings are elusive throughout the literature.

Divorce and unemployment are described as anomic events that reduce an individual’s social integration in society by reducing the strength of their social bonds, therefore increasing the likelihood of suicide (Durkheim 1951[1897]). The loss of a relationship, or losing a job, decreases an individual’s social integration by diminishing the strength and number of social bonds they have with others. These events further prevent society from regulating an individual’s behaviors and desires, which can lead to normlessness. For example, a job provides an individual with a structure, rules, schedule, and a sense of accomplishment. When one loses their job, they may likely lose such sources of regulation, which makes it more difficult for society to regulate their behaviors.

Marriage is generally seen as a protective factor as it further integrates one into society and forms stronger social bonds. Essentially, the greater the intensity of social bonds one possesses in society (marriage, childbearing, employment, church participation, etc.) the less likely one is to suicide. For example, being married and having children depend on an individual makes it less likely for that individual to choose to complete suicide, especially if their family depends on them. Those who lack these bonds (single, live alone, unemployed, etc.) signal a greater potential for feelings of loneliness and lack of social integration, thereby increasing the risk of suicide (Neumayer 2003). These individuals may be more likely to complete suicide as their lack a partner or children that depend on them. Essentially, it is easier to exit existence knowing fewer people will be affected. An understudied variable that may contribute to lower social integration is living alone, which this study includes. For many decades the number of Americans living in single households has increased.

Major depressive disorder, substance use disorders, and serious psychological distress produce modest associations with suicide (Miller et al. 2012). When correlated to the 50-state suicide rate, major depressive disorder, substance use disorders, and serious psychological distress produce a coefficient of 0.4, 0.2, and 0.2 respectively (Miller et al. 2017). While it has been found that around 90 percent of individuals who commit suicide had a mental disorder (Bertolote & Fleischmann 2002), it is also the case that the vast majority (98%) of individuals who experience mental disorders do not die by suicide (Nordentoft et al. 2011). A systematic review of the suicide literature found that depression is the most common psychiatric disorder for those who die by suicide (Hawton et al. 2013). Not having mental health resources to treat such mental illness would further increase one’s likelihood of completing suicide.

Aggregate-Level Studies

Most suicide research is done at the individual-level, however a few studies have been done at the aggregate level. Of these studies, most involve cross-national analyses, mostly focusing on Europe (Stack & Kposowa 2016, Andres 2005, Neumayer 2003, Brainerd 2001, Girard 1993). Fewer studies have focused on the US. Recently, however, there have been aggregate-level studies focusing on firearm ownership and suicides. These state-level studies consistently find that the prevalence of firearms is significantly associated with the state suicide rate (Briggs & Tabarrok 2014, Kposowa 2013, Knopov et al. 2019, Anestis and Houtsma 2018, Vitt et al. 2018, Kposowa et al. 2016, Miller et al. 2007). The main reason for this association is the fact that over half of suicides in the US are completed via firearm.

Other state-level suicide studies have found that political conservatism is associated with higher rates of suicide (Kposowa 2013). Cylus et al. (2014) found that generous state unemployment benefit programs offset the impact of unemployment on suicide rates between 1968 and 2008. Lang (2013) found that when states enact laws requiring insurance coverage to include mental health benefits, the suicide rate decreases significantly by 5 percent. Kaufman et al. (2020) found that increases in the state minimum wage reduced the state suicide rate from 1990 to 2015. Among those with a high school education or less, a one-dollar increase in the minimum wage decreased the suicide rate by about 6 percent. Lastly, a US county-level analysis found that suicide rates were the highest in the most rural counties and lowest in the more urban counties (Rossen et al. 2018). This study also found that a considerable amount of high suicide rate counties are concentrated in the western, mountain, and Appalachian regions, along with Alaska.

Data Challenges

Identifying in advance individuals who will die by suicide thus far has not been possible (Miller et al. 2012). To further complicate this task, suicide mortality is only modestly correlated with the incidence of suicidal acts (Miller et al. 2012). This is in stark comparison to the high correlation between cancer mortality and the incidence of cancer cases. The correlation coefficient of hospital discharges for deliberate self-harm and suicide mortality for 26 states where data are available is 0.20 (Miller et al. 2012) Simply put, actual suicides cannot be predicted effectively even by assessing suicidal acts and self-harm. Additionally, the vast majority of those who become unemployed or divorced do not go on to commit suicide. Lastly, most individuals who experience chronic pain, excessive alcohol consumption, or are of veteran status do not go on to commit suicide either. Arguably one of the best predictors of suicide is firearm ownership, however it is difficult to obtain this data as the CDC's Behavioral Risk Factor Surveillance System (BRFSS) no longer asks about firearm ownership and the General Social Survey lacks statistical power (Lang 2013). Most recently, the FBI's National Instant Criminal Background Check System (NICS) has demonstrated promising results (Lang 2013, Vitt et al. 2018) by tracking the number of new firearm purchase applications, in relation to suicide rates. However, this data is only available at the state level.

Given the increases in suicides every year, it is imperative to identify suicide hotspots in real-time. Unfortunately, data that can provide information about such hotspots lag by several weeks to months. The CDC often releases geographical suicide mortality data after a cluster of suicides has occurred. This might help prevent future suicides, however it does not allow for interventions that attack currently occurring clusters as they spread to vulnerable people. A real-time lens into the current sentiments of individuals in an area would help solve these data issues. These facts make it very difficult to accurately predict who will succumb to suicide. Given the challenges in individual-level suicide prediction, and the dearth of aggregate-level studies done in the US, this project will assess whether a new lens into area-level suicidality, Google searches, may be viable and integrated into the suicide research repertoire.

Leveraging the Power of Google

Considering the substantial and sustained increases in rates of suicide that require a deeper understanding to guide policy and to inform interventions, this research seeks to leverage a potential lens, Google search data, into the deeply private sentiments of individuals. This work rests on the fact that people will ask Google almost anything. In the privacy of their searching, they will, for example, ask whether their husband is gay, if their partner is cheating on them, where confederate flags can be procured, or for anti-Black jokes. We can expect people to ask Google about things that they would not ask their parents, their partner, or their best friend (Stevens-Davidowitz 2017). It follows that Google search data may provide a useful lens into matters that would otherwise be hidden from the epidemiological and sociological inquiry. The current research seeks to assess the potential of Google search data for understanding area-level suicide mortality.

Because Google search data are, for privacy reasons, available only at the aggregate level, the proposed research examines associations between search rates in areas and suicide rates in those same areas. Following on a long sociological tradition of examining such relationships, this inquiry incorporates aggregate-level control variables that have been shown to predict suicide rates in the past. This allows us to ask whether Google search data can add to what has been empirically identified in prior research.

Overwhelmingly, the most popular resource for assessing health-related topics is Google. What makes Google a very attractive data source is the fact that Google has over 90 percent of the market share in online searches in the US, with over 92 percent of Americans being active internet users (Statista 2021). Google is the world's most utilized search engine, handling roughly 2 trillion searches per year, 167 billion searches per month, and 5.5 billion searches per day (Jun et al. 2018), making it the world's largest data source available. Eighty percent of internet users in the US have searched for health-related topics online (Pew 2013). Given these usage statistics, Google search data has provided promising results and has grown rapidly in its use. There has, in fact, been a 20-fold increase in research articles using Google search data from 2009 to 2018 (Arora et al. 2019).

So far, research has shown novel use of Google search data including monitoring the unemployment rate (D'Amuri & Marcucci 2017), predicting the inflow of tourism, tracking home-buying interest, predicting car sales (Choi & Varian 2012, Arora et al. 2019, Brodeur et al. 2021), and predicting the direction of the Dow Jones and the S&P 500 (Hu et al. 2018). A more dismal side of this research emerges when studies assess area-level prejudice. Areas that searched for racially-charged search terms more frequently were also found to be some of the worst-performing districts for Barack Obama during the 2012 presidential elections (Stephens-Davidowitz 2014). Additionally, areas that Googled the racist n-word were associated with an 8.2 percent increase in the all-cause Black mortality rate in that area (Chae et al. 2015). These startling findings show us how aggregate Google searches have the potential to reveal meaningful social patterns and phenomena across geographic areas.

Meaningful social patterns and phenomena are recorded by Google searches as well. For example, when Donald Trump was elected during the 2016 election, searches for “will I be deported” rose precipitously across the US (Chykina & Crabtree 2018). Searches for “will I be deported” also spiked when Trump was inaugurated, and when the administration implemented the “Muslim ban.” In another example, searches looking for off-the-books ways to terminate a pregnancy are higher in states that have passed laws restricting abortions (Stephens- Davidowitz 2017). Searches for terms such as “my mom beat me” and “my dad hit me” increased during the Great Recession and were closely tracked with the unemployment rate (Stephens-Davidowitz 2017). Given that an overwhelming amount of child abuse cases go unreported, and social service agencies were understaffed because of the recession, reports of child abuse failed to observe this disturbing trend. Google searches, on the other hand, were able to shed light on this difficult-to-observe phenomenon. Following this logic, an individual struggling with suicidal thoughts are more than likely to search for topics relating to their stressful situation, and ways to escape their situation via suicide. Therefore, Google search data may have the capacity to reveal to us the unfortunate and painful realities individuals find themselves in.

Currently, public health data are overwhelmingly generated through surveys and reports from numerous agencies that obtain their data from hospitals, local public health departments, and coroners. It can take several months to years to properly collect and analyze these data, which does not allow researchers to assess the data in a timely manner. The recent COVID-19 epidemic highlighted the need for recent and up-to-date data (Drew et al. 2020). What makes Google a novel data source is the fact that the data are collected and reported in real-time, something public health datasets are unable to do (Arora et al. 2019). Since the data are provided in real-time, it is a tremendous improvement over more traditional health surveys such as the CDC's BRFSS which takes over a year to assess and analyze. Therefore, we may be able to assess which areas have more individuals contemplating suicide in real-time by surveying suicide-related search terms. These real-time strategies would give public health initiatives a timely opportunity to implement suicide prevention programs, instead of having to wait months for the latest mortality data.

This Study

This project will test if Google searches allow us to gain further insight into the effects of the social field. As stated earlier, the social field can be measured through the effects produced by changes in the field. This study proposes that effects produced by changes in the social field can be measured by observing what is being searched for on Google. Further information about social fields and measuring their effects can be found in the introduction chapter of this series.

If area-level Google searches can reveal meaningful patterns of thoughts and behaviors, can we leverage this method to assess which areas experience greater levels of suicide ideation, and ultimately suicide? This study argues that this may be the case. The rationale for this association is that the more an area Googles about suicide and suicide-related topics, the more likely residents in such areas are to be considering, contemplating, and learning about suicide. An individual contemplating suicide may be even more inclined to complete once learning about effective methods online. In fact, individuals often seek information regarding suicide and how to effectively kill themselves (Eriksen et al. 2020, Mok et al. 2016, Thornton et al. 2016). An estimated 54 percent of English-language websites that result from suicide-related search terms provide step-by-step information on effective suicide methods (Biddle et al. 2016). Information regarding suicide methods, detailed instructions, and materials are readily available online (Gunnell et al. 2015). Therefore, individuals who may be thinking about suicide, may be curious about suicide methods, or have questions about how to suicide painlessly are likely to Google about it. Google data allows areas where many individuals are making these searches to be revealed in real-time.

This study analyzes the association between the frequency of suicide-related Google search terms and the rate of suicide by geographic area. <u>This study does so and also contributes the following to the literature: 1) makes use of search term composites by combining numerous search terms 2) controls for more covariates than previous research and 3) does the analysis at the smallest unit of analysis available by Google. 4) tests the idea of measuring the effects of social fields using Google search term data.</u>

Past studies have used a small set of search terms which this study argues is not enough to assess the complex and multifaceted phenomena of suicide. This study uses at least 39 different suicide-related search terms and conducts a principal component analysis to examine which terms are Googled most frequently, especially in high suicide risk areas. That is, the results of the principal component analysis will reveal which suicide-related search terms are most common in areas with high rates of suicide. The logic behind this approach is that while many areas may Google for “suicide”, it is far more concerning if an area not only frequently Googles for “suicide”, but searches for “kill myself" and “suicidal” as well. An area that frequently searches for multiple suicide-related terms may be more at risk of having a greater suicide rate. This study argues that by using multiple search terms and creating a composite, it would better capture an area's true inclination towards suicide.

Next, while some studies have observed this association, these studies have not used control variables that are warranted by the extensive literature on this topic. The extensive literature has found numerous risk factors for suicide that go largely ignored by previous studies. This study, on the other hand, makes use of an extensive set of control variables that are informed by the literature. Therefore, this study improves upon previous research by including these theoretically and empirically informed variables. Also, numerous studies examining suicide have tended to not use Google search terms as a variable. These studies may not see the value of Google search term data, or may not be aware of it, whereas this study shows the utility and potential of the data. The paragraphs below discuss previous studies and what they controlled for, and also what this study controls for.

Lastly, the unit of analysis of this study is at the smallest geographic tract available by Google. Previous studies have observed the association between suicide-related searches and the suicide rate at the national level and the state level, however no study has yet done the analysis beneath the state level. This study goes beneath the state level and examines designated market areas of which there are 209 in the US. This level of analysis allows us to gain greater insights into smaller geographic areas, which could be used to efficiently implement public health resources in such areas. Another benefit to studying a smaller scale is that it provides a finer lens for observing social behavior, especially when smaller areas that require attention get lost in state-level analyses.

Several studies have successfully observed the association between suicide-related Google searches and suicide in Japan, South Korea, Taiwan, Germany, among others (Hagihara et al. 2011, Sueki 2012, Song et al. 2014, Chang et al. 2015, Paul et al. 2017). However, very few studies have examined this association in the US, which is surprising given the precipitous rise in suicides over the last two decades (CDC 2021). These studies, their methods, and their findings are described below.

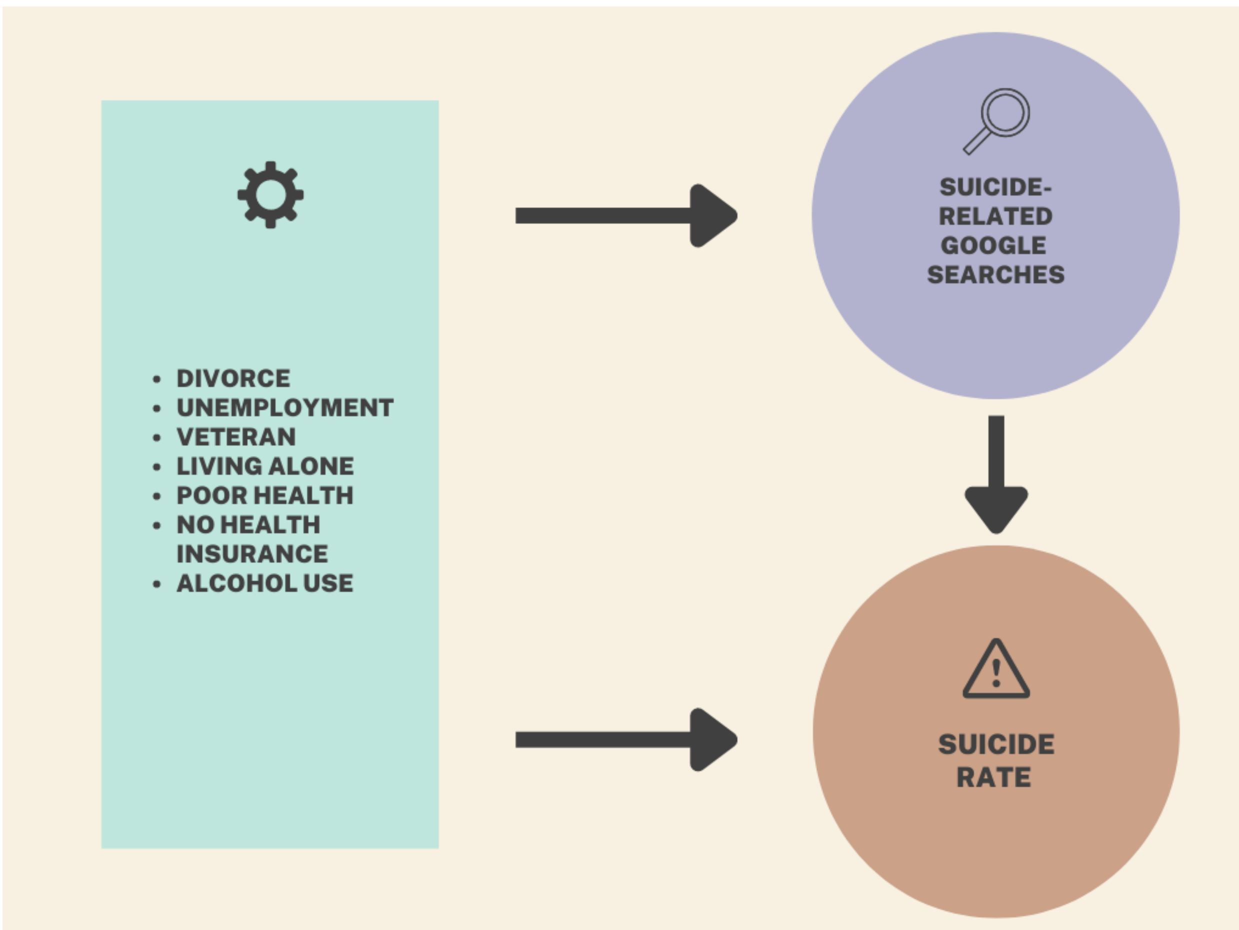
Using Pearson correlation analysis, Gunn & Lester (2013) found that Google searches for “commit suicide”, ”how to suicide”, and “suicide prevention” were positively associated with the 50-state suicide rate in the year 2009. Using an L1-regularization lasso model, Parker et al. (2017) found that life stressor-related Google searches were able to better predict the 50-state suicide rate than alternative models for the year 2015. Using ARIMA models, Capron et al. (2021) found that firearm suicide-related Google searches are associated with the national-level suicide rate from 2004 through 2016. Using Box-Jenkins transfer function models, Lee (2020) found that anxiety, sleep, and unemployment-related terms showed a significant correlation to the national-level suicide rate for 2004 through 2017. Using Box-Jenkins transfer function models, Lee et al. (2021) found the term “gun suicide” to be correlated with the national-level firearm suicide rate from 2004 to 2017. Using mixed model analysis, Adam-Troian & Arciszewski (2020) found that absolutist words such as “completely” and “totally” are predictive of the 50-state suicide rate from 2004 through 2017. Using fixed effects regression, Guerra (2019) found that a Google search term composite comprised of search terms “suicide”, “suicide prevention”, “cyanide”, “suicide hotline”, and “suicidal” are significantly associated with the 50-state suicide rate from 2006 to 2016.

Regarding control variables in these studies, Gun & Lester (2013) only ran Pearson correlations, therefore there were no control variables involved in their analysis. Parker et al. (2017) controlled for state-level unemployment rate and income per capita. Both Capron et al. (2021), Lee (2020), and Lee et al. (2021) controlled for auto-correlational patterns and temporal trends using forecasting models. Adam-Troian & Arciszewski (2020) controlled for between-state structural differences and historical time trends. Guerra (2019) controlled for state-level divorce, unemployment, binge drinking, days mentally unhealthy, handgun ownership, state fixed effects, and yearly fixed effects.

This study goes beyond these previous studies by controlling for percent married, percent divorced, percent unemployed, percent veteran, percent living alone, percent without health insurance, percent reporting poor/fair health, percent reporting heavy alcohol consumption, primary care physicians per 100,000 population, mental health providers per 100,000 population, DMA fixed effects, and yearly fixed effects. Therefore, this analysis will assess if suicide-related search terms are associated with the suicide rate independent of known risk factors. This is the first study to analyze this association at the DMA level.

Figure 1 below illustrates the conceptual map of this study. This study asserts that risk factors such as divorce, being unemployed, being a veteran, living alone, having poor health, heavy alcohol consumption, and not having health insurance increases one's inclination toward suicide. Because individuals often seek information on how to suicide successfully on the internet (Eriksen et al. 2020, Mok et al. 2016, Thornton et al. 2016), these individuals are very likely to Google about suicide and suicide-related topics, which then enables them to commit successful suicide (top arrow). These suicides are then reflected in the area's suicide rate. Additionally, some individuals may experience stressful events such as a divorce or losing their job, and they may not utilize Google, they simply might jump off a high bridge or use a gun (lower arrow). Again, this study will assess the association suicide-related Google searches have on an area's suicide rate.

Figure 1. Conceptual map of the study showing the direction of causality between variables



HYPOTHESIS

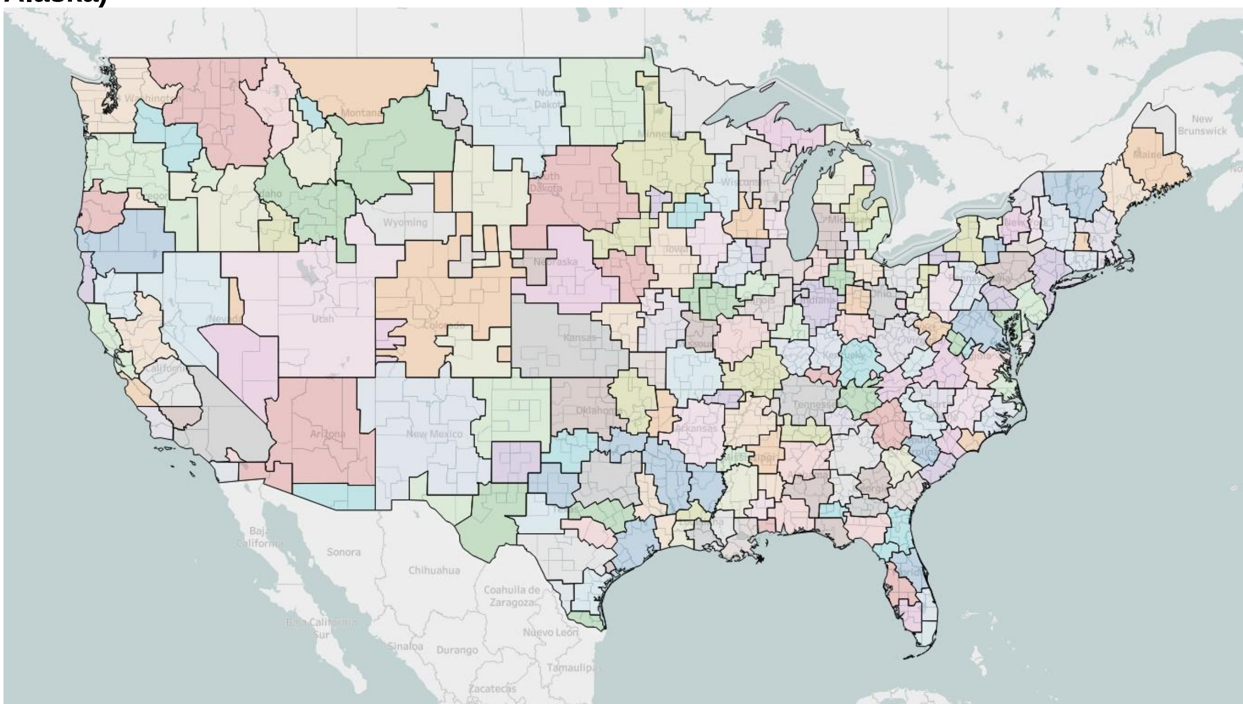
H1: Suicide-related search terms are positively associated with the all-cause suicide rate independent of known area-level risk factors.

DATA

Unit of Analysis

This analysis is done at the Designated Market Area (DMA). There are 210 DMAs in the US. DMAs are designed for media marketing purposes and therefore individuals within these DMAs are expected to share similar characteristics, interests, and profiles. The DMA is the finest geographic level currently provided by Google. One DMA (Fairbanks AK) was dropped from the analysis due to large amounts of missing data. Missing data occurs as a result of low search volume in an area. A DMA map is shown below in figure 2. Alaska and Hawaii are included in the DMAs however they are not pictured in this illustration. DMAs are given different colors to make viewing easier.

Figure 2. Map of Designated Market Areas in the US (Not pictured: Hawaii & Alaska)



Time

This longitudinal analysis is from 2010 to 2016. The year 2010 is chosen because it reflects the widespread adoption of smartphones, which significantly increased the frequency of Google searches. The analysis ends in 2016 as a result of the CDC Compressed Mortality File ending in 2016.

Number of Observations

As there are 209 DMAs and 7 years of data, there are 1463 unique observations.

Independent Variable

Suicide-related search term frequency is obtained from Google Trends. Google Trends is a data tool that provides real-time and past data regarding Google searches. Google Trends provides the frequency that a term has been searched, thereby gauging the popularity of a term over time and in an area. Google Trends data are not user identifiable and are used here at the aggregate DMA level.

Google Trends indicates how frequently a term is searched relative to the total of all other terms searched in the area. This method ensures that areas with the highest populations will not be given the highest scores. The DMA that has the highest frequency of searches relative to the overall total number of searches is assigned a value of 100. The rest of the DMAs are assigned a value of search term frequency that is relative to the DMA with the highest value. This method is consistent with several other studies using Google Trends (Mavragani et al. 2018, Lee 2020, Prado-Roman et al. 2021, Hall et al. 2020).

Suicide-related Google search terms are guided by probabilistic, theoretical, and empirical postulations of key words an individual contemplating suicide is most likely to search for to gain familiarity with suicide methods. This project uses all suicide-related search terms from previous studies as well. This study recognizes that there is no standard method that produces the potential search terms, and acknowledges that other studies have used similar brainstorming procedures. Additionally, this study recognizes that a standardized method for choosing which search terms to obtain would benefit the entire field that uses Google search data.

Several terms had incomplete Google Trends score data. The DMA-level has the least amount of available Google Trends data compared to the state and national-level data. The reason for this is because there are simply less individuals in smaller geographic tracts. Fewer individuals residing in a DMA results in fewer Google queries. For example, as of 2016 the DMA of Fairbanks, located in Alaska, has a population of 5594. There are not enough Google queries in low population DMAs to accurately produce a Google Trends score. Another reason for missing data lies in the wordiness of the search term. While this study found nearly complete data for the search term "suicide", it found substantially less data for the term "I'm having suicidal thoughts". This study finds that at the DMA-level, simple search terms produce less missing data compared to complex and wordy search terms. This is because there are substantially more Google searches for "suicide" than for "I'm feeling suicidal." While this study would argue that the latter term would be much more useful as a suicide predictor, this study must also work with the data that is provided.

An analysis of search term frequency on a national-level confirms that the search term "suicide" is queried substantially more than the terms "kill myself" and "I'm feeling suicidal." Pictured below in figure 3 is the Google Trends frequency scores of all three terms within the last 12 months, with "suicide" in blue, and "kill myself" and "I'm feeling suicidal" pictured in yellow and red respectively. The two largest spikes in searches for "suicide" occurred during the suicides of former 2019 Miss America Cheslie Kryst and country music star Naomi Judd on January 30, 2022 and April 30, respectively.

Figure 3. Google search frequency for "suicide" (blue), "kill myself" (yellow), and "I'm feeling suicidal" (red) from June 27, 2021 to June 27, 2022.

Terms that had no more than 35 percent of its data missing were considered for the analysis. Missing Google Trends score data was imputed using Kalman smoothing. Kalman smoothing is a popular and powerful algorithm that identifies patterns in time series data and creates values based on these patterns (Li et al. 2015).

The suicide-related search terms that met the missing data criteria include the following:

•suicide	•kill myself	•how to die	•ways to die
•suicidal	•suicide hotline	•carbon monoxide	•hang myself
•assisted suicide	•die	•commit suicide	•self harm
•how to suicide	•hydrogen sulfide	•I hate life	•kill yourself
•poison	•suicide by	•suicide death	•death
•want to die	•cyanide	•gun suicide	•carbon monoxide poisoning
•how to overdose	•overdose	•I need help	•kill
•suicidal thoughts	•suicide prevention	•crisis center	•crisis line
•ways to die	•hate life	•help line	•help
•suffer	•suffering		

Because these analyses use an exhaustive amount of search terms, using all search terms in the analyses increases the likelihood of model overfitting, which hinders the model's performance when conducting predictive analyses. In order to prevent model overfitting, and for the sake of model simplicity, this analysis implements a principal component analysis of the search terms of interest. Principal component analysis is a dimensionality-reduction method that reduces the amount of variables in a dataset into a lesser number, while preserving most of the information contained in the data (Abdi & Williams 2010). Principal component analysis identifies variability among correlated variables to produce a small number of variables called components. All search terms of interest were analyzed via principal component analysis for each year from 2010 to 2016. The terms that produced a loading higher than .70 on the first component through at least five years of the analysis were selected to be a part of the search term composite.

Using an exhaustive set of search terms is an improvement from previous search term studies because those studies analyzed DMA-level associations with single terms. For example, Chan (2019) used the search term “Hitler” and ‘kkk” to assess DMA-level intolerance, Hall et al. (2020) used the search term “pollen” to assess seasonal patterns in DMA-area level pollen concentration, Cousins et al. (2020) used a few COVID-19 -related search terms such as “am I sick”, “covid symptoms”, and “covid testing” to assess DMA-level COVID-19 cases, and Chae et al. (2018) used the n-word search term to assess DMA-level racist social climates.

PCA Results

Principal component analysis was conducted for every year of the analysis (2010-2016) using the search terms listed above. The analyses found that the following terms loaded on the first component with at least a .70 score for at least five of the seven years. The terms within this component are the following: **suicide, suicidal, suicide prevention, commit suicide, death, die, kill, kill yourself, how to die, and carbon**

monoxide poisoning. All ten search terms in the composite produce an alpha score of 0.76, which indicates good reliability. Alpha scores below 0.60 are considered unreliable (Vaske et al. 2017). All ten suicide-related search terms were averaged by summing then dividing by ten to produce a suicide-search term composite.

Dependent Variable

The dependent variable is the DMA-level suicide rate per 100,000. Suicide rate data is obtained from the Center for Disease Control's Compressed Mortality File. All-cause suicide is defined by ICD-10 cause codes X60-X84. The suicide rate was then aggregated by each DMA.

The Compressed Mortality File provides suicide counts by county. Counties were merged to their respective DMAs using a DMA index number. Each DMA has several counties within them. County suicide counts were then aggregated by their respective DMA to produce a sum of suicides by DMA. County total population was also aggregated by the respective DMA to produce a DMA total population. DMA suicides were then divided by the total DMA population. This number was then multiplied by 100,000 to produce DMA-level suicide rates per 100,000 persons. County population denominator counts were obtained from the CDC's Population File.

A very small number of suicides in DMAs with very low populations produced extremely high suicide rates which are considered outliers. Outliers that are over 3.5 standard deviations are truncated and given values at the 3.5 standard deviation mark. This is done so that outliers do not significantly affect the regression coefficient results.

Covariates

In order to test if suicide-related Google searches reveal more than what prominent suicide predictors already reveal, this project implements a maximally robust set of control variables.

This project controls for the following:

•Percent married
•Percent divorced
•Percent unemployed
•Percent veteran
•Percent living alone
•Percent without health insurance
•Percent reporting poor/fair health
•Percent reporting heavy alcohol consumption
•Primary care physicians per 100,000 population
•Mental health providers per 100,000 population
•DMA fixed effects
•Yearly fixed effects

Since all control variables are obtained at the county level, they are aggregated to the DMA-level using county population as weights. This was done by multiplying the county variable with its respective county population. This product was then summed with the other products from the same DMA. This sum was then divided by the total population of the DMA to produce an average DMA value that takes into account counties with differing populations. This procedure can also be done by taking the county population and dividing it by the DMA population. The quotient is then multiplied by the respective county variable. This procedure is repeated for each county in a given DMA. Then these products are summed up to produce an average DMA value that takes into account counties with differing populations. This procedure and the one mentioned prior were conducted and achieved identical results.

Percent married, percent divorced, percent unemployed, percent veteran, percent living alone, and percent without health insurance are obtained from the US Census American Communities Survey 5-year estimates and are provided at the county level.

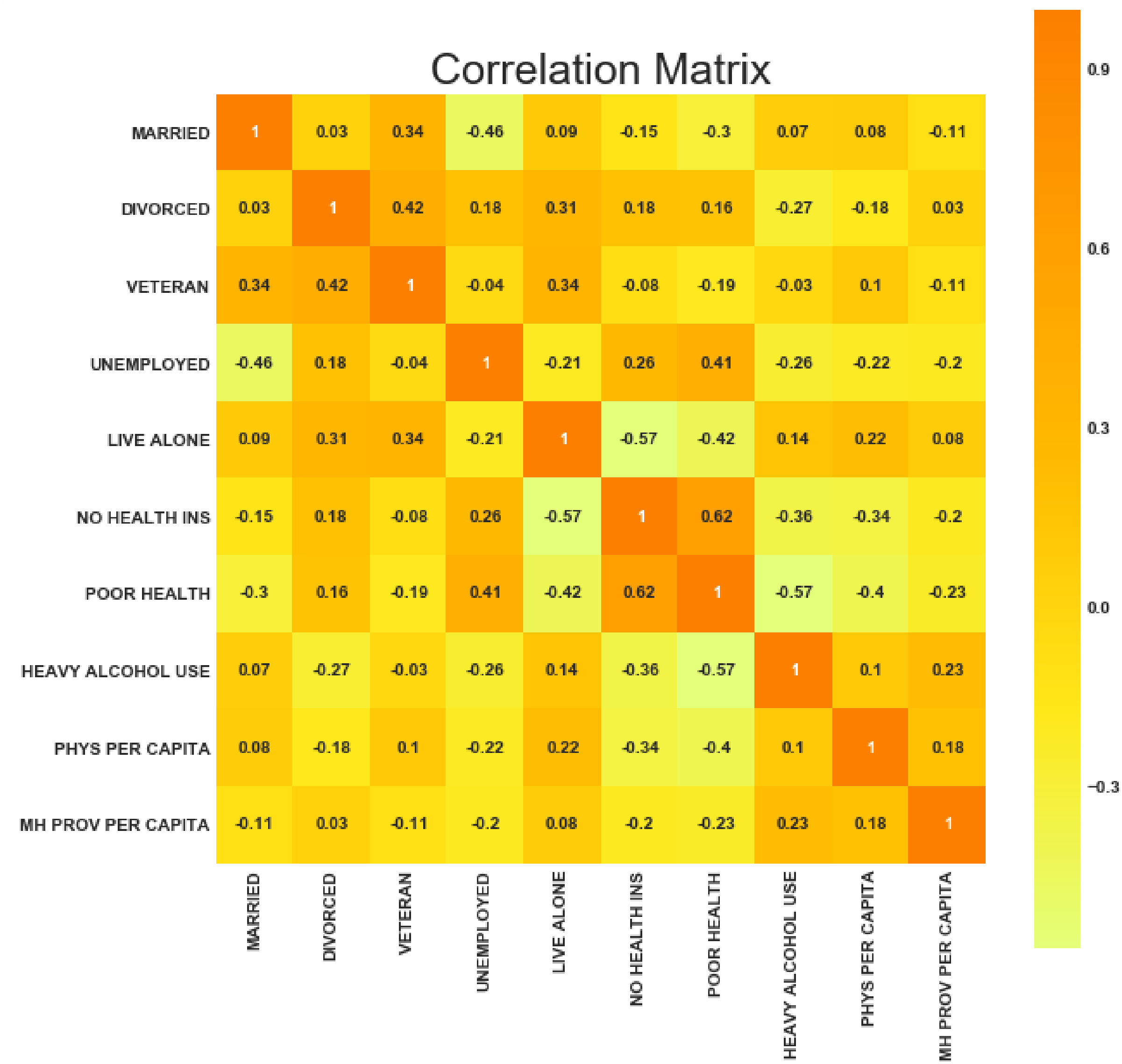
Percent reporting poor/fair health, percent reporting heavy alcohol consumption, primary care physicians per 100,000, and mental health providers per 100,000 are obtained from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) and are provided at the county level. The BRFSS is a nationally representative annual telephone survey conducted by the CDC that assesses health-related risk behaviors throughout the United States. At more than 400,000 adult interviews every year, it is the largest health survey system in the world (CDC 2021).

Data on mental health providers per 100,000 was missing for 2010-2011. Missing value imputation was accomplished via Kalman smoothing. The heavy alcohol consumption variable is defined as consuming more than 4 (women) or 5 (men) alcoholic beverages on a single occasion in the past 30 days, or heavy drinking, defined as drinking more than 1 (women) or 2 (men) drinks per day on average. For the reporting poor/fair health variable, respondents were asked "In general, would you say that your health is excellent, very good, good, fair, or poor?". Respondents answering "poor" and "fair" were coded as poor/fair to create the poor/fair variable.

Multicollinearity

To check for multicollinearity, all variables were assessed for their Variance Inflation Factor (VIF). The poor health variable had a VIF score of 4.5 while the no health insurance variable had a VIF score of 4.4. These were the highest VIF scores out of all the variables. A VIF score above 5 indicates moderate multicollinearity while a score above 10 indicates high multicollinearity. The results of the VIF assessment indicate that multicollinearity is not an issue in these analyses. A correlation matrix of all the variables can be observed below.

Figure 4. Correlation matrix of all covariates in the study



METHODS

The data are analyzed using fixed effects linear regression with robust standard errors clustering for DMA. Two-way fixed effects control for yearly and DMA fixed effects. Model specifications are consistent with previous area-level suicide studies that use two-way fixed effects with robust standard errors clustering by area (Brainerd 2001, Kunce & Anderson 2002, Neumayer 2003, Vitt et al. 2018, Anestis et al. 2017, Stack & Kposowa et al. 2016, Kposowa et al. 2016, Cylus et al. 2014). Model specification was supported via a Hausman test that indicated the appropriateness of using a fixed effects model over the random effects model.

The multivariate fixed effects model to be estimated is as follows:

$$SR_{ij} = \beta ST_{it} + \beta X_{ij} + \gamma_t + \alpha_i + \varepsilon_{ij}$$

The term SR is the suicide rate per 100,000 in DMA *i* at time *t.* The variable of interest, Google search terms, is represented by vector ST. Vector X consists of the following control variables: marriage, divorce, unemployment, veteran status, living alone, not having health insurance, reporting poor health, heavy alcohol consumption, primary care physicians per 100,000, and mental health providers per 100,000. Parameter γ represents time effects that account for unobserved national forces driving the suicide rate. Parameter α represents unobserved DMA-specific fixed effects. The error term is represented by vector ε.

Fixed effect models control for potential omitted unobserved time-invariant area-specific factors that may affect the suicide rate. Time-invariant factors that rarely undergo drastic changes through time are effectively controlled by the model. These factors include culture, climate, geography, elevation, religious composition, gender, and race. DMA-level factors that may contribute to suicide, such as a suicide culture, are controlled for as well.

Numerous studies cite the effect that economic cycles have on suicide rates (Wasserman 1984). Time-specific social currents such as recessions or major events that can affect suicide rates on a national level can provide misleading estimates. In efforts to not confound the estimates with an exogenous variable that may affect suicide trends, the specified model will control for such trends by controlling for yearly effects. It may be the case that a celebrity suicide may spur national increases in the suicide rate, nevertheless such instances are controlled for by the yearly fixed effects. Standard errors are clustered by DMA to address serial correlation which biases standard errors. Analysis is conducted using the fixest package in R.

Model Specification

Certain procedures were undertaken to ensure the best model fit. The dependent variable, the DMA-level suicide rate, was transformed via log, square root, reciprocal, square, and exponential transformation. All transformed dependent variables were assessed for model fit using Bayesian information criterion (BIC) (Neath & Cavanaugh 2012). The model with the lowest BIC, indicating better model fit, was the log and square root model. When compared side to side, the log model performed best. The log model had a slightly better fit than the original non-transformed model. Because the log model had only a slightly better fit than the original model, and to make model interpretation easier, the original model was chosen. Therefore, the dependent variable in this project was not transformed and is the true suicide rate.

Feature Selection

Because of the large number of covariates in the multivariate models, feature selection was conducted to explore the explanatory power of the covariates. Feature selection was done via stepwise selection. Both forward selection and backward elimination procedures were undertaken and produced similar results. Stepwise selection was done to the original, log, and square root models. In all three results, the model with the lowest BIC contained percent living alone and percent reporting poor/fair health. These results are consistent with the models discussed below.

Interaction Effects

All variables were checked for interaction effects. Over forty models were run checking for significant interaction effects. The only interaction effect found was percent living alone and percent divorced, however it was significant at the 0.10 level. No other significant interaction effects were found.

RESULTS

Several models were estimated. First, bivariate linear regressions were run with each variable as the independent variable and the suicide rate as the dependent variable. None of these models included fixed effects. Robust standard errors were clustered at the DMA level. Bivariate models are run to demonstrate each variable's independent relationship with the suicide rate, without controlling for other unobserved confounders. This allows us to verify each variable's association with the suicide rate, however this leaves room for unobserved confounders. Unobserved confounders are controlled for in models after this.

The results of these bivariate models are shown in table 1 below. The suicide search term composite is significant at the 0.001 level in the expected direction, and produces a modest r-squared of 0.07. Additionally, divorce, veteran status, and living alone are significant at the 0.001 level in the expected direction. Marriage is significant at the 0.001 level, however in the unexpected direction. Mental health providers per capita is significant at the 0.001 level in the positive direction. It is unclear what direction mental health providers per capita would be expected as areas high in suicide might attract more mental health resources, however mental health resources in an area should lower suicide rates. Unemployment is significant at the 0.01 level in the unexpected direction. Divorce produces the highest r-squared (0.32), followed by veteran status (0.20), living alone (0.09), and marriage (0.08).

Table 1. Bivariate linear regression. Dependent variable: suicide rate (2010-2016)

Linear Regression (Bivariate)

Dependent Variable: All-Cause Suicide Rate (2010-2016)

	Coefficient	SE	R2
Suicide Search Term Composite	**0.1269*** **	0.0119	0.07
Married	**0.3466*** **	0.0710	0.08
Divorce	**1.689*** **	0.1463	0.32
Unemployment	**-0.2598* **	0.1133	0.02
Veteran	**0.8687*** **	0.1000	0.20
Live Alone	**0.8763*** **	0.1743	0.09
No Health Insurance	0.0354	0.0592	0.00
Alcohol	-0.0165	0.0604	0.00
Poor Health	-0.1522`	0.0791	0.01
Physicians per Capita	-0.0141	0.0076	0.00
Mental Health Providers per Capita	**0.0077*** **	0.0023	0.02
Fixed-Effects:	None		
SE: Clustered	Robust / Clustered by DMA		
Observations	1,463		

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Next, a multivariate linear regression was run with the suicide rate as the dependent variable. This model does not control for any fixed effects. Robust standard errors were clustered at the DMA level. The results of this multivariate model are shown in table 2 below. This model is run to observe the relationship these covariates have with the suicide rate, and to observe which associations are no longer significant as a result of controlling for numerous variables. It is important to note that this model does not control for other unobserved confounders. Unobserved confounders will be accounted for in the next two models.

After for controlling for an exhaustive set of variables, the suicide search term composite is significant at the 0.001 level in the expected direction. This result demonstrates that the suicide search term composite is significantly associated with the suicide rate independent of the major suicide predictors from the literature. Therefore, Google search data adds to what we already know about suicide.

Divorce, veteran status, living alone, and not having health insurance are significant at the 0.001 level in the expected direction. Marriage is significant at the 0.001 level in the unexpected direction. Mental health providers per capita is significant at the 0.001 level. Poor health is significant at the 0.001 level in the unexpected direction.

Table 2. Multivariate linear regression. Dependent variable: suicide rate (2010-2016)

Linear Regression (Multivariate)

Dependent Variable: All-Cause Suicide Rate (2010-2016)

	Coefficient	SE
Suicide Search Term Composite	**0.0899***	0.0147
Married	**0.2420***	0.0545
Divorce	**1.1253***	0.1355
Unemployment	-0.0973	0.0758
Veteran	**0.3467***	0.0818
Live Alone	**0.4261***	0.2083
No Health Insurance	**0.1752***	0.0485
Alcohol	-0.0031	0.0697
Poor Health	**-0.1696***	0.0804
Physicians per Capita	0.0003	0.0075
Mental Health Providers per Capita	**0.0090***	0.0013
Fixed-Effects:	None	
SE:	Robust / Clustered by DMA	
Observations	1,463	
R2	0.52	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

The models above demonstrate that the data is valid and most of the variables produce expected results. The models also demonstrate that the variable of interest, the suicide search term composite, is significantly associated with the suicide rate for the period of 2010 to 2016.

The following models include yearly and DMA fixed effects. It is important to note that the fixed effects effectively control for time-invariant factors. Several of the variables in the models can be described as time-invariant, such as percent marriage, divorce, veteran status, living alone, unemployment, not having health insurance, alcohol consumption, reporting poor health, and physicians per capita. For example, over the course of time, the percent of individuals who are veterans does not undergo drastic changes; it remains relatively stable. The fixed effects in the model will omit the effects these time-invariant factors have on the suicide rate and as a result, the results produced by these time invariant factors should be evaluated keeping these facts in mind. In fact, a descriptive analysis of the national-level trends of these variables shows that all of them are time-invariant with the exception of mental health practitioners per capita.

Table 3 below describes the national-level trend of each covariate for the year 2010 and 2016. Between these seven years, marriage decreased from 50.2 percent to 48.1percent. Divorce increased from 10.5 percent to 11 percent. The unemployment rate decreased slightly from 7.9 percent to 7.4 percent. Veteran status decreased from 9.9 percent to 8 percent. Living alone stayed the same at 10.2 percent. Not having health insurance decreased from 21.1 percent to 20.6 percent. Heavy alcohol consumption increased from 15 percent to 17.5 percent. Reporting poor health increased from 16.1 to 16.5 percent. Physicians per capita increased from 69 to 73 per 100,000. The largest change observed were mental health practitioners per capita which increased significantly from 115 to 164 per 100,000. From these trends, we can observe that changes in these covariates, with the exception of mental health practitioners per capita, are very likely to be captured, and therefore controlled for, by the fixed effects models discussed below. Therefore, several of the variables that are significant in Tables 1 and 2 will cease to remain significant because the fixed effects have essentially controlled them.

Table 3. Descriptive statistics of covariates showing little change between 2010 and 2016

Descriptive Statistics

US National-Level	**2010**	**2016**
Married	50.20%	48.10%
Divorced	10.50%	11.00%
Unemployed	7.90%	7.40%
Veteran	9.90%	8.00%
Live Alone	10.20%	10.20%
No Health Insurance	21.10%	20.60%
Alcohol	15%	17.50%
Poor Health	16.10%	16.50%
Physicians per Capita	69	73
Mental Health Prov per Capita	115	164

Source: US Census Bureau & CDC BRFSS

Bivariate linear regressions were run with each variable as the independent variable and the suicide rate as the dependent variable. These models include both yearly and DMA-level fixed effects. Robust standard errors were clustered at the DMA level. These bivariate models are run to demonstrate each variable's independent relationship with the suicide rate, while controlling for unobserved confounders. This allows us to verify each variable's association with the suicide rate.

The results of these bivariate models are shown in table 4 below. The suicide search term composite is significant at the 0.05 level in the expected direction. This demonstrates that the suicide search term composite remains significant after controlling for unobserved confounders.

All previously significant variables are no longer significant after implementing the DMA-level and yearly fixed effects. The poor health variable emerges as significant at the 0.05 level in the unexpected direction. The living alone variable produces the largest r-squared (0.01), followed by poor health (0.005), and the suicide search term composite (0.005).

Table 4. Two-way fixed effects bivariate linear regression. Dependent variable: suicide rate (2010-2016)

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Bivariate)

Dependent Variable: All-Cause Suicide Rate (2010-2016)

	Coefficient	SE	R2
Suicide Search Term Composite	**0.0375***	0.0157	0.005
Married	-0.0247	0.2585	0
Divorce	0.2113	0.3800	0
Unemployment	0.0652	0.1623	0
Veteran	0.0307	0.4035	0
Live Alone	1.0865`	0.5600	0.01
No Health Insurance	-0.0816	0.0634	0.002
Alcohol	-0.0750	0.0981	0.001
Poor Health	**-0.1789***	0.0704	0.005
Physicians per Capita	0.0136	0.0083	0.003
Mental Health Providers per Capita	0.0009	0.0013	0
Fixed-Effects:	Year & DMA		
SE: Clustered	Robust / Clustered by DMA		
Observations	1,463		

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Lastly, a multivariate linear regression was run with the suicide rate as the dependent variable. These models include both yearly and DMA fixed effects. Robust standard errors are clustered at the DMA level. The results of this multivariate model are shown in table 5 below.

This model is run to observe the relationship the covariates have with the suicide rate, and to see which associations are no longer significant as a result of controlling for the covariates. Additionally, this model allows us to observe which variables remain significant after controlling for unobserved confounders.

The variable of interest, the suicide search term composite, is significant at the 0.01 level producing a coefficient of 0.04. Controlling for all confounders, a ten-point increase in suicide-related search interest is associated with a 0.4 point rise in the suicide rate. For an area of one million people, this equates to about 4 extra deaths, or roughly 100 years of potential life lost. Years of potential life lost is calculated by subtracting the age of suicide by the standard life expectancy of 75, then taking the conservative assumption that all deaths occurred around the age of 50 (CDC 2018). Living alone is significant at the 0.05 level and produces a considerable coefficient of 1.39 in the expected direction. The poor health variable is significant at the 0.01 level in the unexpected direction.

The suicide search term composite is significant in the expected direction in every single model discussed so far. The search term composite demonstrates its reliability by remaining significant while undergoing several models while controlling for some of the most prominent predictors of suicide from the literature.

Table 5. Two-way fixed effects multivariate linear regression. Dependent variable: suicide rate (2010-2016)

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Multivariate)

Dependent Variable: All-Cause Suicide Rate (2010-2016)

	Coefficient	SE
Suicide Search Term Composite	**0.0409***	0.0160
Married	0.3031	0.2878
Divorce	0.1216	0.4814
Unemployment	0.1649	0.1565
Veteran	0.0866	0.4242
Live Alone	**1.3995***	0.5708
No Health Insurance	-0.1081	0.0612
Alcohol	-0.0773	0.0869
Poor Health	**-0.2011****	0.0747
Physicians per Capita	0.0138	0.0077
Mental Health Providers per Capita	0.0008	0.0013
Fixed-Effects:	Year & DMA	
SE:	Robust / Clustered by DMA	
Observations	1,463	
R2	0.03	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

The suicide search term composite is significant in the expected direction in all models listed above. This demonstrates that it is a reliable construct that is capable of capturing beyond what is already known by the literature. Since it remains significant after controlling for the major predictors of suicide, it demonstrates that Google searches can add to what we already know about suicide. After showing that it remains significant after controlling for unobserved confounders, it shows that it is significant independently of all other factors and therefore it can be argued that Google searches provide us with another way of observing an area’s inclination towards suicide in and of itself.

Lagged Model

While the models discussed above assess the variables and the suicide rate on a yearly basis, this approach can bring up questions of causality. Therefore, all independent variables were lagged by one year and regressed on next year's suicide rate. For example, the suicide search term composite of 2015 was regressed on the 2016 suicide rate. To conduct this analysis, the 2009 entries for all independent variables were obtained. Every bivariate model was run with the suicide rate as the dependent variable, and the one-year lag of each covariate as the independent variable, while controlling for the one-year lag of the dependent variable, the suicide rate. No variables were significant in these bivariate analyses. Next, a multivariate model was run with the suicide rate as the dependent variable and the one-year lags of all covariates, including the one-year lag of the dependent variable, the suicide rate. No variables were significant in the multivariate analysis. The fact that there are no significant findings indicates to the reduced ability in using the one-year lag of suicidal search terms in attempting to predict next year's suicide rates. While the year-by-year analyses provide promising results, this study cannot assert that the suicide search term composite has any significant predictive power.

Robustness Check

It is possible that suicide-related Google search terms may be confounding with an unobserved aspect of an area rather than a specific phenomenon related to suicide. To test this possibility, we examined other possible health outcomes, cancer mortality and heart attack mortality, as dependent variables. If the suicide-related Google search terms are significantly associated with such outcomes, then this study cannot argue on the reliability of Google data and its association with the suicide rate.

Cancer mortality and heart attack mortality rate data was obtained from the CDC compressed mortality file. Cancer mortality is defined by ICD-10 codes C00-C97. Heart attack mortality is defined by ICD-10 codes I210-I219. Cancer and heart attack mortality rates were aggregated by DMA and merged to the final dataset.

The suicide search term composite was tested in a bivariate fixed effects model controlling for DMA and yearly fixed effects. The suicide search term composite did not have an association with cancer mortality ($t = 1.22$). The suicide search term composite did not have an association with heart attack mortality ($t = -0.36$).

Therefore, these analyses strengthen the assertion that Google search term data may reveal profound social currents that dispose individuals towards succumbing to the most personal of all choices, and that such associations are not spurious in nature.

DISCUSSION

The discussion section summarizes the findings from the result sections, focusing on the suicide search term composite, and arguing that Google search data can add to what is already known from the literature. It touches on the strengths of implementing Google search composites opposed to using single search terms. It then mentions another novel finding, how living alone can be a risk factor for suicide. The discussion finishes by citing the importance of using Google search data and other internet sources that can provide researchers with rich information on human attitudes, behaviors, and inclinations.

The suicide search term composite is significantly associated with the DMA suicide rate. Therefore, the extent that a DMA Googles the following terms: suicide, suicidal, suicide prevention, commit suicide, death, die, kill, kill yourself, how to die, and carbon monoxide poisoning, is significantly associated with that DMA's suicide rate over the period of 2010 through 2016. Just a ten-point increase in the search term frequency for the suicide search term composite is associated with 0.04 more suicide deaths for 100,000 population. For an area of one million people, this equates to about 4 extra deaths, or roughly 100 years of potential life lost. This study provides some evidence for the utility of Google search terms to measure suicide-related interest at the area level. To the best of our knowledge, while other studies have done so at the national and state-level, this is the first study to assess the association between the frequency of suicide-related searches and the suicide rate at the DMA-level.

It is imperative to note that this association remained significant even after controlling for the popular suicide predictors from the sociological, epidemiological, psychiatric, and suicidology literature. Therefore, Google search terms may be able to provide us with additional information in addition to what is already known about area-level suicide risk factors. After controlling for unobserved confounders via the fixed effects model, the suicide search composite remained significant indicating that it is significant independently of all other factors and therefore it can be argued that Google searches provide us with another way of observing an area's inclination towards suicide in and of itself.

Using principal component analysis aiming to create a suicide search composite was successful in this study. The principal component analysis found that the following terms loaded reliably on a factor: suicide, suicidal, suicide prevention, commit suicide, death, die, kill, kill yourself, how to die, and carbon monoxide poisoning. The principal component analysis analyzed forty different suicide search terms, which is more than previous studies have explored. This analysis gives some credence to the idea that search term composites may assist in assessing social phenomena better than single terms. The logic behind this is that an area might report a spike in searches for “suicide” because of a local suicide on the news, however if an area reports a spike in searches for “suicide”, “commit suicide”, and “carbon monoxide poisoning”, this may be a more comprehensive method of observing an area’s true inclination towards suicide. Therefore, Google search composites may provide a truer and more holistic lens into an area’s innermost thoughts and inclinations.

No association was observed for the one-year lag of the variable of interest, indicating that the suicide search term composite does not possess predictive power over a year of lag. A reason for this null finding may be the fact that the yearly structure of the data prevents a more refined analysis about timing. For example, it is very unlikely for a suicidal individual to Google about suicide in 2010 and then complete the suicide in 2011. However, it is much more likely to for a suicidal individual to Google about suicide in May and then complete the suicide in June. A more informative analysis would have a monthly structure. A monthly data structure would allow us to observe the association between the monthly lag in suicide searches and completed suicides. However, monthly suicide data at the county or DMA-level may be difficult to procure from the CDC. A more viable method to test month lags would be at the state-level.

To summarize, the results of the models discussed above demonstrate that the variable of interest, the suicide search term composite, is robust and reliable throughout the entire series of models, even after controlling for the most popular predictors of suicide from the literature. Additionally, the variable of interest remains significant after controlling for unobserved confounding, giving credence to the idea that Google searches can add to what we already know about suicide. As previous studies using

Google search terms have been able to assess numerous health outcomes, including herpes, plague, cancer, whooping cough, influenza, asthma, Lyme disease, syphilis, HIV, depression, obesity, drug use, and many more, (Mavragani & Ochoa, 2019, Arora et al. 2019, Jun et al. 2018, Brodeur et al. 2021), this study has successfully used a suicide search term composite to assess the suicide rate at the DMA-level over the period of 2010 to 2016.

Living Alone

What is most interesting is that, besides the variable of interest, most of the covariates in the analyses can be argued to be time-invariant. That is, within these DMAs, these variables do not undergo drastic changes throughout the years, and are in fact, rather stable through time. However, a trend that has increased substantially throughout the last few decades is the fact that more Americans are living alone than ever before. Twenty-eight percent of households, or one in seven Americans, are living alone (Chamie 2021). The living alone variable has shown interesting results in these models. In the bivariate linear regression model (Figure 1), it is significant at the 0.001 level and produces a coefficient of 0.87. In the multivariate linear regression model (Figure 2), it is significant at the 0.05 level and produces a coefficient of 0.42. Every one percent increase in the population living alone is associated with a 0.42 increase in the suicide rate. For an area of one million people, this equates to roughly four suicides per year, or about 120 years of potential life lost. In the fixed effects bivariate model (Figure 3), living alone is significant at the .10 level and produces a coefficient of 1.08. Then in the fixed effects multivariate model (Figure 4), it is significant at the 0.05 level and produces a coefficient of 1.39, which for an area of one million people, we can expect an additional 14 suicides per year, or 420 years of potential life lost.

These findings should be investigated further as Americans increasingly find themselves in single person households. From a Durkheimian perspective, living alone diminishes one's social integration in society by reducing one's ties to immediate family and a significant other. While research has found that single people report a greater number of social connections (Sarkisian & Gerstel 2016), it is possible that the quality of these connections may not protect against suicide as well as marriage and childbearing has. However, the data does not identify what type of household the person who completed suicide lived in. Future research should investigate what type of household individuals who have completed suicide come from. From these findings, a possible hypothesis would be that individuals who complete suicide are more likely to come from single-person households.

Google and internet data

Regarding Google searches, these findings call for further research into our online digital traces and how they may provide researchers with clues about area-level health. While causality is difficult to ascertain from these analyses, the significant associations found warrant further investigation of how area-level Google search behavior is tied to health and mortality outcomes. This study demonstrates that what areas Google about is significantly associated with the suicide rate of such areas. In other words, in areas where suicide is prevalent, we observe area residents Googling about these very events as they occur, before they occur, and after they occur. In a sense, observing what an area Googles about frequently gives researchers a glimpse of the social conditions, or social climate, of such areas.

These findings give credence to the idea that social climates, such as those described by Mark Hatzenbueher (social attitudes and social norms of an area), can be observed using internet-based measures. Given that roughly 90 percent of Americans use Google as their search engine, and nearly 93 percent of Americans have internet access via computer or smartphone (Pew Research Center 2021), Google data gives researchers a rare glimpse of almost an entire area's thoughts and behaviors. The extent that Google is used by so many individuals and so frequently throughout their day, provides us an almost unparalleled source of information on social life. This study gives credence to the idea that an area's interests, thoughts, and behaviors are reflected in Google search data. These data may assist in assessing further concepts such as culture, social trends, prejudicial attitudes, and more.

In light of Google search data producing promising results throughout numerous fields, the internet provides additional sources of social information such as Twitter, Facebook, and Reddit. These sources provide rich amounts of data as they oftentimes provide detailed descriptions of individual experiences and stories. Using internet data in research has increased precipitously in the last decade, which has brought us an era of computational social science. The internet offers an unprecedented wealth of text, image, and audio data produced on numerous sites and applications, which provides us with the opportunity to catalog human history on an unprecedented scale (Mohr et al. 2020). New forms of data may allow us to identify patterns that could expand the range of research questions social scientists may ask for years to come (Mohr et al. 2020).

LIMITATIONS

Not all Google searches for suicide-related content are necessarily made by people contemplating suicide or looking for suicide methods. Nonetheless, an area conducting a more than average search volume for suicide-related content can be said to be reflecting the overall social environment of such area. For example, Googling for

“COVID symptoms” does not necessarily mean one has COVID, but it does reflect a situation that is influenced by the surrounding social environment. The argument that Google searches for suicide-related content are not necessarily made by suicidal people would be supported by non-significant findings, however that is not the case. This study has shown that the more that areas search the terms suicide, suicidal, suicide prevention, commit suicide, death, die, kill, kill yourself, how to die, and carbon monoxide poisoning, the higher their suicide rates were revealed to be.

Religiosity is a sociological construct that should have a negative effect on suicide, and therefore protect against suicide (Durkheim 1951[1897]). However, data on area-level religiosity can only be found at the state-level through the General Social Survey and therefore data at the DMA-level is not available. Using Google Trends data to measure area-level religiosity has not been thoroughly tested and is a potentially new area for exploration. Thus far, there has been one study investigating religiosity via Google search data with promising results (Yeung 2019).

Also, given that roughly half of all suicides are completed via firearm, a major suicide risk factor is firearm ownership. However, data on this is elusive and only available at the state-level. The General Social Survey stopped asking about firearm ownership two decades ago. Currently, the most reliable data on firearm ownership is from the FBI’s National Instant Criminal Background Check System (NICS) and is only available at the state level. Several studies have shown the NICS’s viability in assessing the suicide rate at the state-level (Lang 2013, Vitt et al. 2018). However, obtaining reliable firearm ownership data below the state-level remains a challenge.

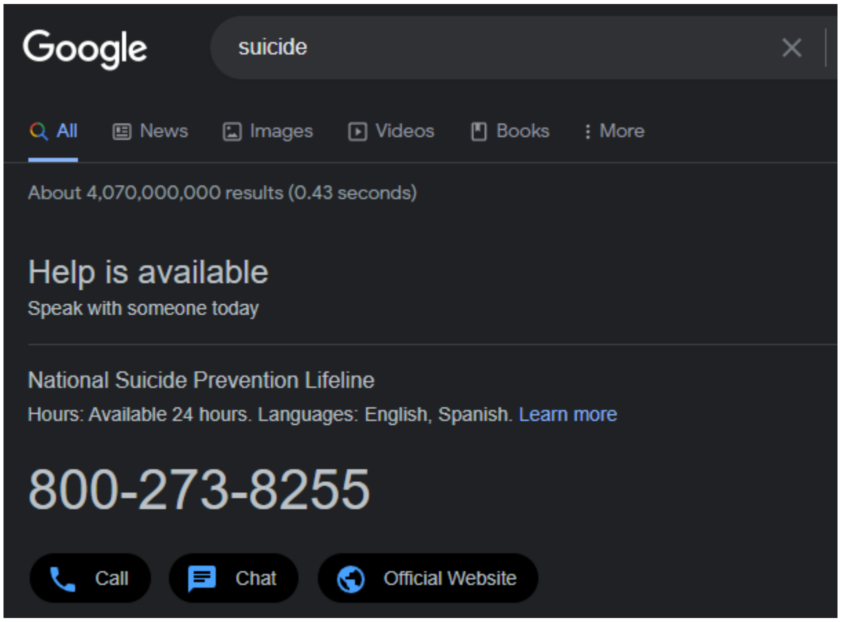
This analysis does not make any claims about causation, nor can causality be established from this analysis. Instead, this analysis explores a new lens of observing the social world that is reflected in an area's Google search behavior. This analysis identifies suicidal search term data as a potential leading measure of suicidal behavior. Google search term data may be more viable and more closely associated with the suicide rate than other established measures. The development of this methodology assists in the development of further analyses and theory which would develop pathways to observe suicidal inclinations and prevent suicides in real-time.

POLICY IMPLICATIONS

DMAs that frequently Google for these terms should be watched closely by the CDC. Areas that frequently Google information about suicide may indicate multiple area residents are considering suicide. Knowledge on which areas are most at risk informs policy and health initiatives, such as where to allocate suicide prevention resources. Thankfully these data are available in real-time and therefore DMAs that experience a rise in suicide-related search term frequency can be identified and targeted swiftly with prevention efforts. For example, the systematic screening of suicide-related search terms, which reflects real-time search behaviors in an area, can assist in monitoring area-level suicide-interest and suicide risk.

Currently, a cost-effective and practical initiative to help prevent suicides has Google editing their search algorithm in a way that delivers prompts to users who query suicide-related terms. These prompts include information on where to get help with suicide ideation. An example of such can be observed below. A good research design could explore the difference in suicide rates before and after such initiatives are rolled out to test the effectiveness of such.

Figure 5. Prompt issued by Google when searching for suicidal topics



ACADEMIC CONTRIBUTIONS

This project advances epidemiology and suicidology research by implementing an innovative methodology that is seldom used. Additionally, this project advances the emerging field of digital epidemiology, whose studies show that our online digital traces leave behind clues about our health. Previous studies have validated this methodology at the national and state-level for suicide, however no previous studies have done these analyses at the DMA-level. This study contributes to the study of suicide by being the first to do these analyses at the DMA-level, which is the smallest geographic tract available in Google Trends.

This project advances our theoretical understanding of the social field, and how effects of the social field can be measured by using Google search data. This study argues that effects produced by changes in the social field can be measured by observing what is being searched for on Google. As previously noted, changes in the social field, such as the election of Donald Trump, led to a spike in searches for “will I be deported”. Similarly, changes in the social field, such as a divorce, can drive individuals to consider suicide, and their sentiments are often captured in Google searches. Since this study found the suicide search term composite to be robust throughout a series of models, these findings demonstrate that Google search data is capable of capturing an individual’s innermost thoughts, attitudes, and inclinations. Furthermore, this source of data has the potential to compliment what is already known about suicide risk-factors. The findings of these analyses support Google search term data as an additional lens into social behavior.

This project advances sociological research in that it introduces an emerging and verified methodology to its field. Sociologists have not identified innovative and sociologically interesting approaches to suicide that are available in other disciplines (Wray et al. 2011). Current sociological research lacks up-to-date methods and instead focuses heavily on surveys and interviews. Furthermore, this project revives a long sociological tradition of studying suicide. Since 1980, of the 30,000 academic articles on suicide, only 1.3 percent were categorized as sociological (Wray et al. 2011). Given the dearth of recent sociological research on suicide, this project fills many gaps in the methods of studying suicide.

CONCLUSION

The suicide search term composite used in this study is significantly associated with the DMA all-cause suicide rate between 2010 and 2016. This analysis gives support to analyses using Google Trends as a viable source of data that captures the behaviors, attitudes, and inclinations of a population. In this instance, Google Trends is shown to better capture the variance in suicide rates than traditional data sources and methods.

Association between Drug-Related Google Searches and the Drug Overdose Rate

Abstract: Fatal drug overdose is a major public health and social problem in the US. Google search term data may help researchers understand variations in area-level fatal drug overdose rates. Drug-related Google search interest is highly correlated with area-level fatal drug overdose rates. This analysis tests whether drug-related Google searches are associated with area-level fatal drug overdose rates from 2010 through 2016 using two-way fixed effects as a means of controlling unmeasured confounding. The analysis controls for major overdose predictors including area-level educational attainment (high school and below), service occupation employment, percent white, poverty, unemployment, poor health, and heavy alcohol use. Results demonstrate drug-related search term interest is associated with area-level fatal drug overdose rates, and shows promising results for prediction.

INTRODUCTION

Drug overdose is increasingly becoming a public health issue in the US. While several predictors from the literature provide reliable assessments of area-level mortality, these assessments are useful after overdoses have occurred. This study proposes using a new lens of observing social phenomena, real-time Google search term data, to understand variations in area-level fatal drug overdose rates. Google provides information on how popular search terms are in areas over time. Therefore, **it is possible to test if area-level drug-related Google queries are associated with the fatal drug overdose rate**. And, if this association exists, then it opens the possibility to assess drug-related Google search behavior in real time.

Furthermore, this project will test if Google searches allow us to gain further insight into the effects of the social field. The social field is an integral part of social field theory, which examines patterns of interaction and behavior between individuals and the surrounding environments where they move about (Burnes & Bargal 2017). Social field theory is used to explain how social forces can have an influence on individuals even if they are not in direct contact with one another. Social forces in the social field act on individuals and influence their thoughts, feelings, and behaviors even though there is no specific observable contact between individuals (Mohr. et al. 2020, 117). The social field can be measured through the effects produced by changes in the field. For example, when US states ban abortion, which is a change in the social field, it has been observed that these states Google more frequently for off-the-book ways of obtaining an abortion (Stephens-Davidowitz 2017).

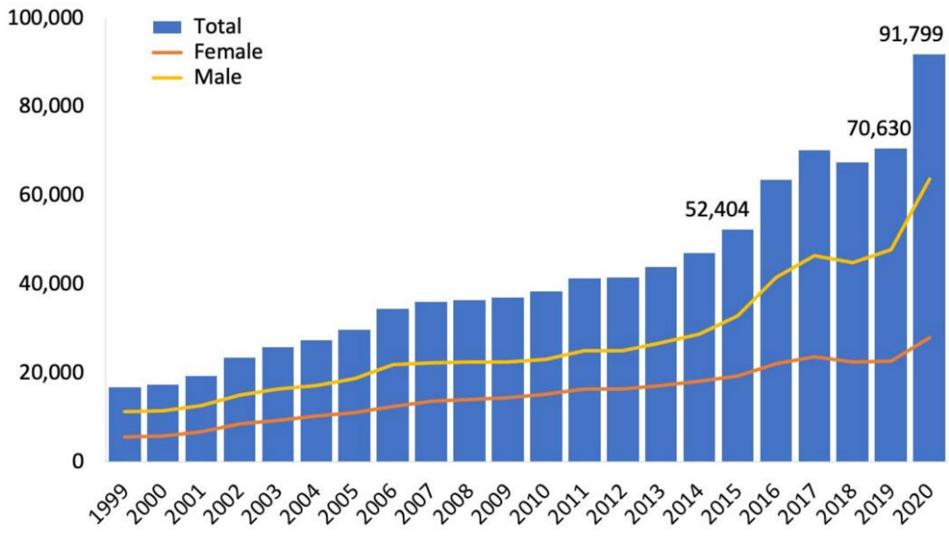
In another example, Google searches for “will I be deported” rose precipitously after the election of Donald Trump (Chykina & Crabtree 2018). According to social field theory, the field can only be measured through the effect of changes in the social field. Therefore, the effect of Donald Trump being elected can be observed by the rise in searches for “will I be deported.” These examples demonstrate that Google search data has the capacity to reveal to us the unfortunate and painful realities many individuals find themselves in. As mentioned above, this project tests if area-level drug-related Google searches are associated with fatal drug overdose rates using longitudinal data spanning seven years.

This paper begins with a review of the overdose problem in the US and the major predictors. It then describes the utility of using Google data and why it is a good and novel source of data. This section is followed by a review of previous studies that have tested the association between drug-related Google searches and fatal drug overdoses. That section is followed by the major contributions of this project, followed by the data, methods, results, and discussion sections.

Drug Overdose is a Major Public Health Problem

There has been a 300 percent increase in fatal drug overdoses in the US within the last 20 years (CDC 2021). In 1999, nearly 20,000 Americans died from a drug overdose; by 2019, 70,000 Americans have died from the same cause. This equates to roughly one million deaths attributable to drug overdose within the last 20 years (CDC 2021). The number of annual deaths due to drug overdose is visualized in the graph below (National Institute of Drug Abuse 2022).

Figure 1. Number of annual deaths attributable to drug overdose in the US from 1999 to 2020



The rise in drug overdoses has led to a statistical occurrence that has not happened in 100 years; for the first time in a century, the average life expectancy in the US declined. The last time life expectancy declined in the US was in 1915-1918 as a result of military deaths in World War I and the 1918 influenza pandemic (DeWeerdt 2019). The US is the only industrialized nation to experience this decline in life expectancy. These statistics have led the US government to declare this crisis a public health emergency. US states and the federal government have spent billions of dollars attempting to ameliorate this crisis which worsens every year.

Of the nearly 92,000 drug overdose deaths in the year 2020, <u>nearly eighty percent of them involved the use of an opioid</u> (National Institute of Drug Abuse 2022). Psychostimulants, mostly methamphetamine, accounted for roughly ten percent of drug overdose deaths. Cocaine accounted for roughly four percent of drug overdose deaths. Benzodiazepines accounted for roughly three percent and antidepressants accounted for roughly three percent as well. These statistics highlight the dangers and prevalence of opioids in these deaths. For these reasons, the rise of drug overdose deaths is often referred to as the opioid crisis. This opioid crisis can be described as an epidemic of substance use disorders and drug overdoses for which opioids are largely responsible.

Overdoses involving psychostimulants, cocaine, and benzodiazepines have also increased in the last two decades. Additionally, deaths from alcohol-related diseases have also increased substantially (Case & Deaton 2015). Therefore, the rise in overdoses for all types of major drugs indicates a larger societal trend in drug use that requires a deeper sociological understanding. This project will assess all known risk factors, and then assess whether a newer lens into people's innermost thoughts can help us further understand this alarming trend.

Predictors of Overdose

According to epidemiological studies, there are numerous risk factors for opioid overdose death. Males are two times as likely to die from an overdose compared to females. Males between the ages of 30 to 39 experience the highest risk of overdose death, followed by males between the ages of 45 to 49, 50 to 54, 40 to 44, and 25 to 29 respectively. Males who experience the least risk of overdose death are those ages 65 and above (CDC WISQARS). Females between the ages of 50 to 54 experience the highest risk of overdose death, followed by females between the ages of 45 to 49, 55 to 59, 40 to 44, and 35 to 39. Females who experience the least risk of overdose death are those ages 65 and above (CDC WISQARS).

By race, an overwhelming number of whites have died from drug overdoses. In the last ten years, roughly 400,000 non-Hispanic whites have died from opioid drug overdoses. However, in the same period, roughly 50,000 blacks have died from drug overdoses, and about 35,000 Hispanics have died from the same causes (CDC WISQARS). These figures must be understood in the context where roughly 60 percent of the US population is non-Hispanic white, 13 percent is black, and 18 percent is Hispanic. When death rates are taken into consideration, the racial group with the highest rate of drug overdose in 2019 were Native Americans (~26.5), followed by whites (~24.7), blacks (~20.6), Hispanics (~10.5), and Asians (~3.0).

Strong attention to overdose deaths and alcohol-related deaths came into the national spotlight after Case and Deaton’s (2015) article detailing the rising levels of mortality, especially for whites during midlife. Case and Deaton (2015) have described these deaths as “deaths of despair.” Their research has found that these deaths have been primarily driven by individuals with a high school degree or less (Case & Deaton 2021). Individuals with some college (less than a BA) experienced little change in mortality rates, while individuals with a Bachelor’s degree or more experienced a reduction in mortality rates. Therefore, a large part of these deaths are being driven by individuals who lack a college education or never graduated high school. The association between educational attainment and health has been well-documented and multiple mechanisms have been identified including education's influence on participating in healthier lifestyles, obtaining high-paying occupations, providing access to health care, avoiding financial hardships, and developing social capital (Montez et al. 2019).

In addition to lacking a college education, research indicates that areas experiencing economic insecurity are at a heightened risk of overdose. Large-scale deindustrialization as a result of international trade policies has left many American cities and communities with minimal occupational opportunities, thereby exacerbating economic insecurity, especially among those who are less educated (Dean & Kimmel 2019). This is consistent with research by Monnat et al. (2019) that found that US counties with the highest drug overdose rates are characterized by higher levels of economic disadvantage and a greater proportion of blue-collar and service sector employment. Research by Knapp et al. (2019) found that US counties experiencing economic insecurity had higher rates of deaths of despair and all-cause midlife mortality. They found that counties in the highest tertile of economic insecurity had 41 percent higher mortality rates than counties with stable low economic insecurity.

Economic insecurity is often linked with high levels of unemployment. Unemployment is hypothesized to reduce or eliminate opportunity structures that are conducive to upward mobility, such as opportunities for work or education, the creation of pro-social ties, and the availability of suitable marriage partners (Wilson 2011). The social and economic deprivation produced by unemployment contributes to the deterioration of family and community structures, producing social isolation and a sense of despair that is tied to drug use (Kwan et al. 2008).

A systematic review by Nagelhout & Hummel (2017) found that unemployment is associated with drug use. Additional studies have linked unemployment and poor economic conditions to drug use and drug overdose deaths (Nosrati et al. 2019, Ghertner & Groves 2018). Qualitative research on economically insecure areas shows that the decline of high-paying and secure employment has manifested in a collective sense of despair, family breakdown, community breakdown, and a rise in substance abuse (Quinones 2015). These findings are consistent with recent research that found that economic disadvantage, unemployment, and the deterioration of social and opportunity structures are tied to overdose deaths (Dean & Kimmel 2019, Nagelhout & Hummel 2017, Monnat et al. 2019).

Although deaths of despair have been the topic of numerous recent studies, despair remains an ill-defined concept and does not have an agreed-upon definition (Harper et al. 2021). Case & Deaton (2020) acknowledge that the function of the despair concept is descriptive, and not explanatory. In addition to economic despair, unemployment, and the deterioration of family and social structures, Case et al. (2020) have identified dramatic increases in self-reported pain, especially for those in mid-life, to be constituent of this despair. Since the construct of despair is a complex social phenomenon attributable to numerous factors, many of which can have synergistic effects, this study takes a broad definition approach to the construct of despair.

Current efforts to ameliorate this crisis involve limiting access to prescription opioids (Davis et al. 2019), imposing stricter rules and penalties on pain management clinics (Dowell et al 2016, Popovici et al. 2018), and harm reduction measures such as access to Naloxone (Rees et al. 2019). Efforts to identify hotspots often rely on post-mortality data obtained from the CDC and hospital discharge data. However this data lags by several weeks to months. Given the urgency of this topic, it is imperative to identify overdose hotspots in real-time. This study tests a potential real-time lens into the deeply private sentiments that are constitutive of this despair.

Leveraging the Power of Google

Considering the substantial and sustained increases in rates of overdose that require a deeper understanding to guide policy and to inform interventions, this research seeks to leverage a potential lens, Google search data, into the deeply private sentiments of individuals. This work rests on the fact that people will ask Google almost anything. In the privacy of their searching, they will, for example, ask whether their husband is gay, if their partner is cheating on them, where confederate flags can be procured, or for anti-Black jokes. We can expect people to ask Google about things that they would not ask their parents, their partner, or their best friend (Stevens-Davidowitz 2017). It follows that Google search data may provide a useful lens into matters that would otherwise be hidden from the epidemiological and sociological inquiry. The current research seeks to assess the potential of Google search data for understanding area-level drug overdose mortality.

Because Google search data are, for privacy reasons, available only at the aggregate level, the proposed research examines associations between search rates in areas and overdose rates in those same areas. This study incorporates aggregate-level control variables that have been shown to predict fatal drug overdose rates in the past. This allows us to ask whether Google search data can add to what has been empirically identified in prior research.

Overwhelmingly, the most popular resource for assessing health-related topics is Google. What makes Google a very attractive data source is the fact that Google has over 90 percent of the market share in online searches in the US, with over 92 percent of Americans being active internet users (Statista 2021). Google is the world's most utilized search engine, handling roughly 2 trillion searches per year, 167 billion searches per month, and 5.5 billion searches per day (Jun et al. 2018), making it the world's largest data source available. Eighty percent of internet users in the US have searched for health-related topics online (Pew 2013). Given these usage statistics, Google search data has provided promising results to the extent that there has been a 20-fold increase in research articles using Google search data from 2009 to 2018 (Arora et al. 2019).

So far, research has shown novel use of Google search data including monitoring the unemployment rate (D'Amuri & Marcucci 2017), predicting the inflow of tourism, tracking home-buying interest, predicting car sales (Choi & Varian 2012, Arora et al. 2019, Brodeur et al. 2021), and predicting the direction of the Dow Jones and the S&P 500 (Hu et al. 2018). A more dismal side of this research emerges when studies assess area-level prejudice. Areas that searched for racially-charged search terms more frequently were also found to be some of the worst-performing districts for Barack Obama during the 2012 presidential elections (Stephens-Davidowitz 2014). Additionally, areas that Googled the racist n-word were associated with an 8.2 percent increase in the all-cause Black mortality rate in that area (Chae et al. 2015). These startling findings show us how aggregate Google searches have the potential to reveal meaningful social patterns and phenomena across geographic areas.

Meaningful social patterns and phenomena are recorded by Google searches. For example, when Donald Trump was elected during the 2016 election, searches for “will I be deported” rose precipitously across the US (Chykina & Crabtree 2018). Searches for “will I be deported” also spiked when Trump was inaugurated, and when the administration implemented the “Muslim ban.” In another example, searches looking for off-the-books ways to terminate a pregnancy are higher in states that have passed laws restricting abortions (Stephens- Davidowitz 2017). Searches for terms such as “my mom beat me” and “my dad hit me” increased during the Great Recession and were closely tracked with the unemployment rate (Stephens-Davidowitz 2017). Given that an overwhelming amount of child abuse cases go unreported, and social service agencies were understaffed because of the recession, reports of child abuse failed to observe this disturbing trend. Google searches, on the other hand, were able to shed light on this difficult-to-observe phenomenon. Following this logic, an individual struggling with drug use is more than likely to search for topics relating to their stressful situation. Therefore, Google search data may have the capacity to reveal to us the unfortunate and painful realities individuals find themselves in.

Currently, public health data are overwhelmingly generated through surveys and reports from numerous agencies that obtain their data from hospitals, local public health departments, and coroners. It can take several years to properly collect and analyze these data, which does not allow researchers to assess the data in a timely manner. The recent COVID-19 epidemic highlighted the need for recent and up-to-date data (Drew et al. 2020). What makes Google a novel data source is the fact that the data are collected and reported in real-time, something public health datasets are unable to do (Arora et al. 2019). Since the data are provided in real-time, it is a tremendous improvement over more traditional health data such as CDC mortality and hospital discharge data. Therefore, we may be able to assess which areas have more citizens experiencing situations that would enable them to search for drug-related content in real-time by surveying drug-related search terms. These real-time strategies would give public health officials a timely opportunity to implement overdose prevention programs, instead of having to wait for the latest mortality data.

This Study

This project will test if Google searches allow us to gain further insight into the effects of the social field. As stated earlier, the social field can be measured through the effects produced by changes in the field. This study proposes that effects produced by changes in the social field can be measured by observing what is being searched for on Google. Further information about social fields and measuring their effects can be found in the introduction chapter of this series.

If area-level Google searches can reveal meaningful patterns of thoughts and behaviors, can we leverage this method to assess which areas experience greater levels of drug interest, drug use, and ultimately drug overdose? This study argues that this may be the case.

An individual may be more likely to Google about drug-related topics, given the situation they may be experiencing. An individual may be more likely to Google about their addiction, drug cravings, or drug-related topics rather than asking their partner, friends, and peers (Ayers et al. 2013). Furthermore, an individual may be likely to Google information about a drug, the best method of doing the drug, the effects of certain doses, and where to procure drugs. In fact, the internet has rich information that is readily available regarding the use of illicit drugs, their effects, and methods of use

(Davey et al. 2012). Following this logic, if an area has numerous individuals engaged in drug use, several of them are likely to query topics related to such drug use. Additionally, an individual can learn that one of their loved ones are engaged in drug use, and then query about the drug and its dangers. In this example, the person searching on Google is not a drug user themselves, however their online searches are constitutive of their surrounding social networks and social surroundings, which have influenced them to Google. Therefore, the search frequency of drug-related searches becomes an artifact of the surrounding social structure. The frequency of these queries would be recorded and can then be assessed along with the fatal drug overdose rate of such areas.

Therefore, this study analyzes the association between the frequency of drug-related Google search terms and the rate of fatal drug overdose by geographic area. <u>This study does so and also contributes the following to the literature: 1) makes use of search term composites by combining numerous search terms 2) controls for more covariates than previous research and 3) does the analysis at the smallest unit of analysis available by Google.</u>
4) tests the idea of measuring the effects of social fields using Google search term data. Past studies have used a small set of search terms which this study argues is not enough to assess the complex and multifaceted phenomena of drug overdose. This study uses at least 64 different drug-related search terms and conducts a principal component analysis to examine which terms can be used while preserving the richness of the data.

The logic behind this approach is that while many areas may Google for “overdose”, it is far more concerning if an area not only frequently Googles for “overdose”, but searches for “opioids" and “heroin” as well. An area that frequently searches for multiple drug-related terms may be more at risk of having a greater drug overdose rate. This study argues that by using multiple search terms and creating a composite, it would better capture an area's true inclination towards drug use, and ultimately fatal overdose.

Next, while some studies have observed this association, these studies have not used control variables that are warranted by the extensive literature on this topic. The extensive literature has found numerous risk factors for fatal drug overdose that go largely ignored by previous studies. This study, on the other hand, makes use of an extensive set of control variables that are informed by the literature. Therefore, this study improves on previous research by including these theoretically and empirically informed variables. Also, numerous studies examining fatal drug overdose have tended to not use Google search terms as a variable. These studies may not see the value of Google search term data, or may not be aware of it, whereas this study shows the utility and potential of the Google search term data. The paragraphs below discuss previous studies and what they controlled for, and also what this study controls for.

Lastly, the unit of analysis of this study is at the smallest geographic tract available by Google. Previous studies have observed the association between drug-related searches and fatal drug overdoses at the national-level and the state-level, however no study has yet done the analysis beneath the state-level. This study goes beneath the state level and examines designated market areas of which there are 209 in the US. This level of analysis allows us to gain greater insights into smaller geographic areas, which could be used to efficiently implement public health resources in such areas. Another benefit to studying a smaller scale is that it provides a finer lens for observing social behavior, especially when smaller areas that require attention get lost in state-level analyses.

Thus far there have been very few studies that have attempted to test if there is an association between drug-related Google searches and fatal drug overdoses. Using an Extremely Random Forest Machine Learning model, Campo et al. (2020) found that the following Google terms were associated with the 50-state fatal drug overdose rate from 2004 to 2017: Narcan, heroin, opioid, Vivitrol, suboxone, naloxone, perc, overdose, methadone, and withdrawal. Using linear regression, Mukherjee et al. (2020) found that Google searches for drug names such as “heroin”, “fentanyl”, “oxycodone”, and “cocaine” were useful in forecasting weekly drug overdoses in the state of Connecticut from 2012 to 2018. Using a hierarchical regression model, Arendt (2020) found that the Google search term “fentanyl” was associated with the 50-state fatal drug overdose rate from 2004 to 2017.

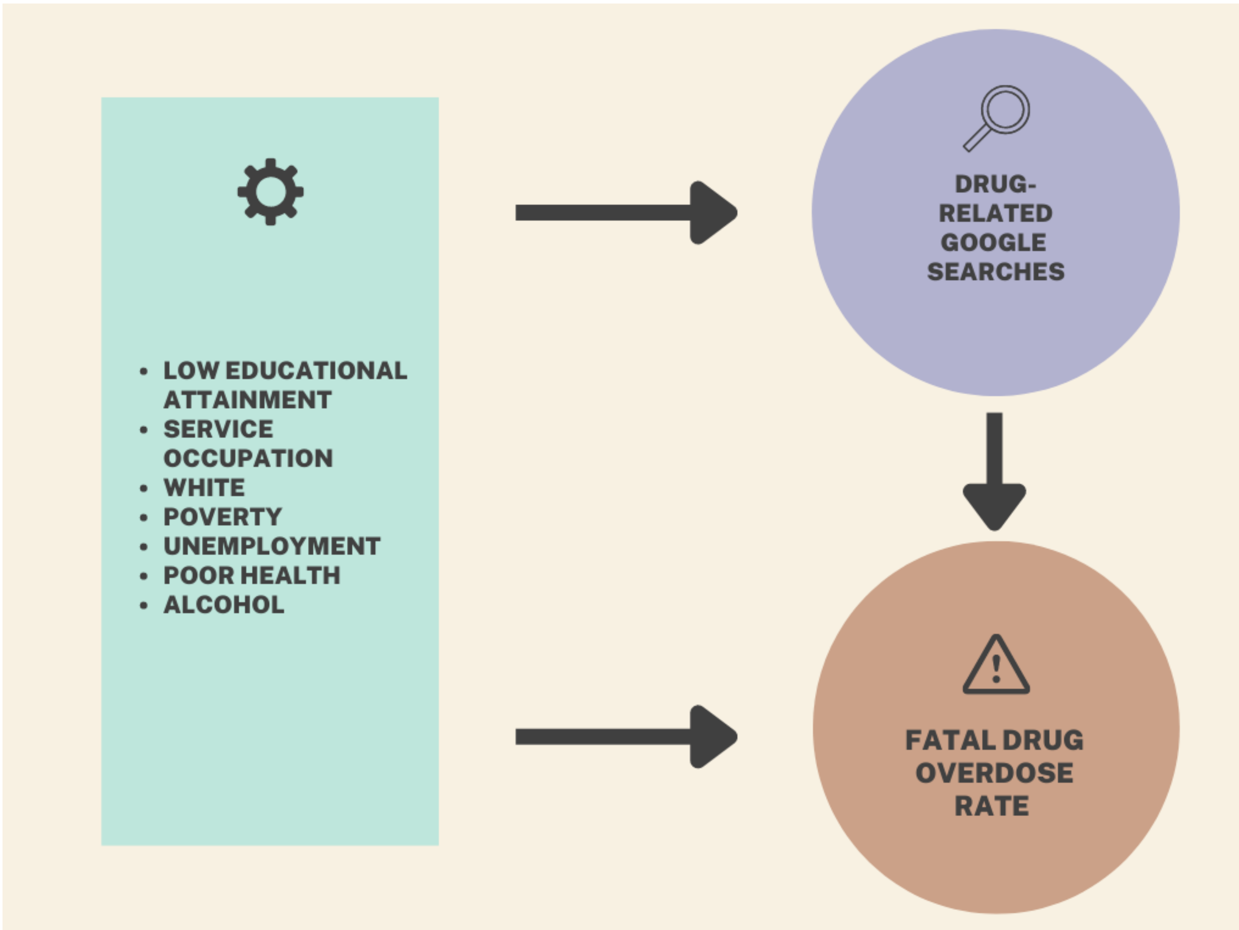
Regarding control variables in these studies, Campo et al (2020) did not include any control variables. Mukherjee et al. (2020) controlled for historical weather data including maximum temperature, minimum temperature, average precipitation, and average snowfall. They utilized these controls citing the effects of weather on mental well-being and happiness. They hypothesized that rainy, snowy, and cold weather would increase the number of overdoses. Arendt (2020) controlled for state-level median age, sex ratio, percent with a college degree, per capita income, percent white, and population density.

This study goes beyond these previous studies by controlling for percent educational attainment (high school and less), percent of workforce employed in service occupations, percent white, percent living in poverty, percent unemployed, percent reporting poor/fair health, percent reporting heavy alcohol consumption, DMA fixed effects, and yearly fixed effects. Given the dearth of knowledge on this subject, this study will explore if there is an association between drug-related Google searches and the fatal drug overdose rate at the Designated Market Area (DMA) level. This is the first study to analyze this association at the DMA level.

Figure 2 below illustrates the conceptual map of this study. This study asserts that risk factors such as low educational attainment, being employed in the service sector, being white, poverty, being unemployed, poor health, and consuming excessive amounts of alcohol increases one's inclination toward having a fatal drug overdose (bottom arrow).

Because individuals regularly seek drug information (Kuehn 2011), ways to procure illicit drugs, and seek methods to use drugs online (Miller & Sonderlund 2010), these individuals are very likely to use Google in these endeavors (top arrow). Drug-related searches would then enable individuals to engage in drug use, which then can in turn lead to a fatal drug overdose (middle arrow). This study will assess the association drug-related Google searches have on an area's fatal drug overdose rate.

Figure 2. Conceptual map of the study showing the direction of causality between variables



HYPOTHESIS

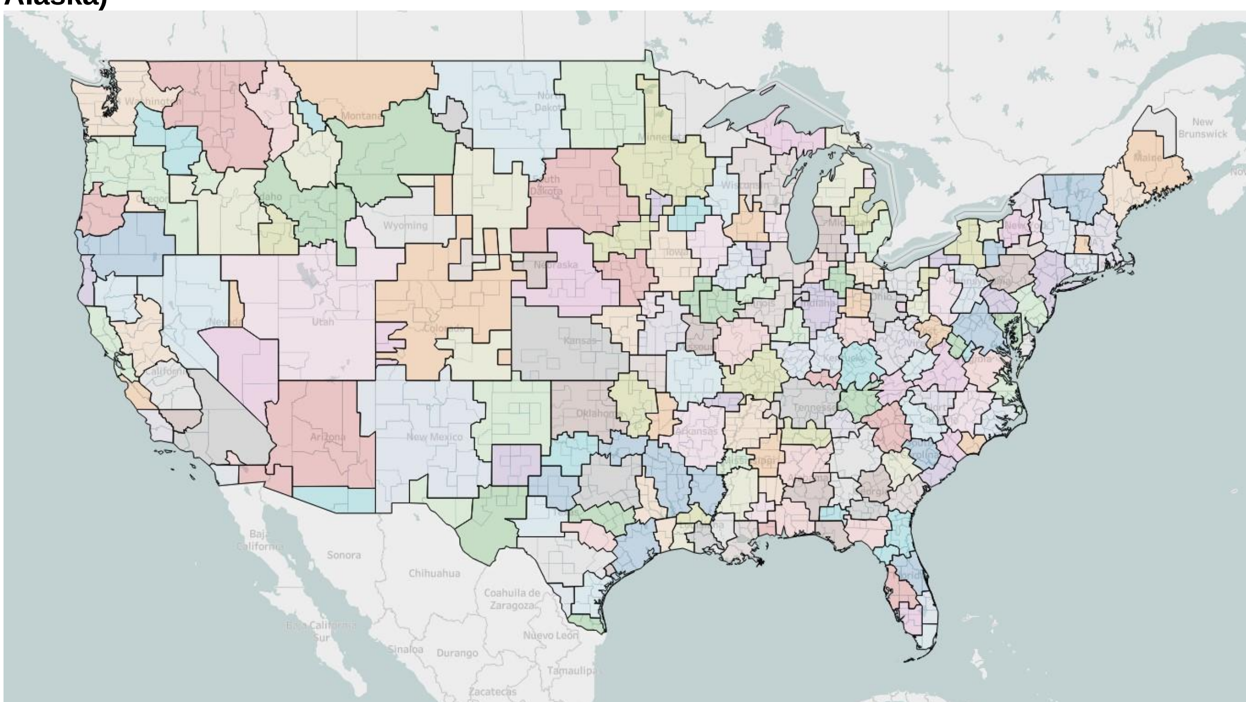
H1: Drug-related search terms are positively associated with the fatal drug overdose rate independent of known area-level risk factors.

DATA

Unit of Analysis

This analysis is done at the Designated Market Area (DMA). There are 210 DMAs in the US. DMAs are designed for media marketing purposes and therefore individuals within these DMAs are expected to share similar characteristics, interests, and profiles. The DMA is the finest geographic level currently provided by Google. One DMA (Fairbanks AK) was dropped from the analysis due to large amounts of missing data. Missing data occurs as a result of low search volume in an area. A DMA map is shown below provided by MapBox. Alaska and Hawaii are included in the DMAs however they are not pictured in this illustration. DMAs are given different colors to make viewing easier.

Figure 3. Map of Designated Market Areas in the US (Not pictured: Hawaii & Alaska)



Time

This longitudinal analysis is from 2010 to 2016. The year 2010 is chosen because it reflects the widespread adoption of smartphones, which significantly increased the frequency of Google searches. The analysis ends in 2016 as a result of the CDC Compressed Mortality File ending in 2016.

Number of Observations

As there are 209 DMAs and 7 years of data, there are 1463 unique observations.

Independent variable

Drug-related search term frequency is obtained from Google Trends. Google Trends is a data tool that provides real-time and past data regarding Google searches. Google Trends provides the frequency that a term has been searched, thereby gauging the popularity of such a term over time and in an area. Google Trends data are not user identifiable and are used here at the aggregate DMA level.

Google Trends indicates how frequently a term is searched relative to the total of all other terms searched in the area. This method ensures that areas with the highest populations will not be given the highest scores. The DMA that has the highest frequency of searches relative to the overall total number of searches is assigned a value of 100. The rest of the DMAs are assigned a value of search term frequency that is relative to the DMA with the highest value.

This method is consistent with several other studies using Google Trends (Mavragani et al. 2018, Lee 2020, Prado-Roman et al. 2021, Hall et al. 2020).

Drug-related Google search terms are guided by probabilistic, theoretical, and empirical postulations of key words an individual facing drug cravings, seeking drug-related help, seeking drug dependency medication, or overdose reversal medication, are most likely to search for. This project uses all drug-related search terms from previous studies as well. This study recognizes that there is no standard method that produces the potential search terms, and acknowledges that other studies have used similar brainstorming procedures. Additionally, this study recognizes that a standardized method for choosing which search terms to obtain would benefit the entire field that uses Google search data.

Several terms had incomplete Google Trends score data. The DMA-level has the least amount of available Google Trends data compared to the state and national-level data. The reason for this is because there are simply less individuals in smaller geographic tracts. Fewer individuals residing in a DMA results in fewer Google queries. For example, as of 2016 the DMA of Fairbanks, located in Alaska, has a population of 5594. There are not enough Google queries in low population DMAs to accurately produce a Google Trends score. Another reason for missing data lies in the wordiness of the search term. While this study found nearly complete data for the search term “heroin”, it found substantially less data for the term “heroin addiction hotline”. This study finds that at the DMA-level, simple search terms produce less missing data compared to complex and wordy search terms. This is because there are substantially more Google searches for “heroin” than for “heroin addiction hotline.” While this study would argue that the latter term would be much more useful as an overdose predictor, this study must also work with the data that is provided.

An analysis of search term frequency on a national-level confirms that the search term “heroin” is queried substantially more than the terms “heroin addiction” and “heroin addiction hotline.” Pictured below are the Google Trends frequency scores of all three terms within the last 12 months, with “heroin” in blue, and “heroin addiction” and “heroin addiction hotline” pictured in red and yellow respectively.

Figure 4. Google search frequency for “heroin” (blue), “heroin addiction” (red), and “heroin addiction hotline” (yellow) from June 27, 2021 to June 27, 2022.

Terms that had no more than 35 percent of its data missing were considered for the analysis. Missing Google Trends score data was imputed using Kalman smoothing. Kalman smoothing is a popular and powerful algorithm that identifies patterns in time series data and creates values based on these patterns (Li et al. 2015).

The drug-related search terms that met the missing data criteria include the following:

•AA	•cocaine	•codeine	•crack
•addiction	•alcohol	•despair	•hopeless
•dilaudid	•oxy	•demerol	•meperidine
•distress	•drug abuse	•addicted	•drug use
•drug	•drugs	•hangover	•help line
•heroin	•overdose	•hydrocodone	•marijuana
•meth	•methamphetamine	•desoxyn	•naltrexone
•morphine	•methadone	•naloxone	•strain

•narcan	•opiate	•oxycodone	•oxycontin
•opioid	•adderall	•addict	•hydromorphone
•percocet	•rehab	•sober	•sobriety
•stress	•stressful	•substance abuse	•suffer
•suffering	•tramadol	•pain killers	•unemployed
•unemployment	•whiskey	•xanax	•withdrawal
•vivitrol	•fentanyl	•suboxone	•perc
•opana	•oxymorphone	•NA	•vicodin
•alprazolam	•back pain	•I want to get high	•get high

Because these analyses use an exhaustive amount of search terms, using all search terms in the analyses increases the likelihood of model overfitting, which hinders the model's performance when conducting predictive analyses. In order to prevent model overfitting, and for the sake of model simplicity, this analysis implements a principal component analysis of the search terms of interest. Principal component analysis is a dimensionality-reduction method that reduces the amount of variables in a dataset into a lesser number, while preserving most of the information contained in the data (Abdi & Williams 2010). Principal component analysis identifies variability among correlated variables to produce a small number of variables called components. All search terms of interest were analyzed via principal component analysis for each year from 2010 to 2016. The terms that produced a loading higher than .70 on the first component throughout all of the years of the analysis were selected to be a part of the search term composite.

Using an exhaustive set of search terms is an improvement from previous search term studies because those studies analyzed DMA-level associations with single terms. For example, Chan (2019) used the search term “Hitler” and ‘kkk” to assess DMA-level intolerance, Hall et al. (2020) used the search term “pollen” to assess seasonal patterns in DMA-area level pollen concentration, Cousins et al. (2020) used a few COVID-19 -related search terms such as “am I sick”, “covid symptoms”, and “covid testing” to assess DMA-level COVID-19 cases, and Chae et al. (2018) used the n-word search term to assess DMA-level racist social climates.

PCA Results

The principal component analysis found that the following terms consistently loaded heavily on a single component (loading higher than .70 on the first component throughout all of the years): **heroin, naloxone, naltrexone, Narcan, overdose, Vicodin, vivitrol, opiate, opioid.** Naloxone (brand name: Narcan), is a potentially life-saving medication designed to help reverse the effects of an opioid overdose in minutes (Narcan.com). Naltrexone (brand name: Vivitrol) is a medication used to treat both alcohol use disorder (AUD) and opioid use disorder (OUD). Vicodin is a highly addictive opiate-based medication used to relieve moderate to severe pain. Heroin, opiate, and opioid refer to opiate-based drugs or medications used either to treat pain or used recreationally. All nine terms of the composite are related to opioids, which is consistent with the fact that nearly eighty percent of fatal drug overdoses are opioid-related (National Institute of Drug Abuse 2022). Given that the majority of fatal drug overdoses involve opioids, this composite is appropriate for the analysis. All nine search terms in the composite produce an alpha score of 0.70, which indicates good reliability. Alpha scores below 0.60 are considered unreliable (Vaske et al. 2017). All nine opioid-related search terms were averaged by summing then dividing by nine to produce a drug-search term composite.

Dependent Variable

The dependent variable is the DMA-level fatal drug overdose rate per 100,000. Drug overdose rate data is obtained from the Center for Disease Control's Compressed Mortality File. Drug overdose is defined by ICD-10 cause codes X40-X44 and Y10-Y14. The fatal drug overdose rate is then aggregated by each DMA.

The Compressed Mortality File provides fatal overdose counts by county. Counties were merged to their respective DMAs using a DMA index number. Each DMA has several counties within them. County fatal overdose counts were then aggregated by their respective DMA to produce a sum of fatal overdoses by DMA. County total population was also aggregated by the respective DMA to produce a DMA total population. DMA fatal overdoses were then divided by the total DMA population. This number was then multiplied by 100,000 to produce DMA-level fatal drug overdose rates per 100,000 persons. County population denominator counts were obtained from the CDC's Population File.

A very small number of drug overdoses in DMAs with very low populations produced extremely high drug overdose rates which are considered extreme outliers. Outliers that are over 3.5 standard deviations are truncated and given values at the 3.5 standard deviation mark. This is done so that outliers do not significantly affect the regression coefficient results.

Covariates

In order to test if drug-related Google searches reveal more than what prominent overdose predictors already reveal, this project implements a maximally robust set of control variables.

This project controls for the following covariates:

•Percent educational attainment (high school and below)
•Percent of workforce employed in service occupations
•Percent white
•Percent living in poverty
•Percent unemployed
•Percent reporting poor/fair health
•Percent reporting heavy alcohol consumption
•DMA fixed effects
•Yearly fixed effects

Since all control variables are obtained at the county level, they are aggregated to the DMA-level using county population as weights. This was done by multiplying the county variable with its respective county population. This product was then summed with the other products from the same DMA. This sum was then divided by the total population of the DMA to produce an average DMA value that takes into account counties with differing populations. This procedure can also be done by taking the county population and dividing it by the DMA population. The quotient is then multiplied with the respective county variable. This procedure is repeated for each county in a given DMA. Then these products are summed up to produce an average DMA value that takes into account counties with differing populations. This procedure and the one mentioned prior were conducted and achieved identical results.

Percent educational attainment, percent of workforce employed in service occupations, percent white, percent unemployed, and percent without health insurance are obtained from the US Census American Communities Survey 5-year estimates and are provided at the county level.

The educational attainment variable is defined as the percentage of those over the age of 25 who have equal to or less than a high school diploma or its equivalent. This variable is consistent with the research of Case & Deaton (2021) who found that drug overdose deaths have been primarily driven by individuals with a high school degree or less. The service occupation variable is defined as the percentage of those who are working employed in a service sector occupation that includes food preparation, retail positions, and service-related occupations. This variable is consistent with research by Monnat et al. (2019) that found US counties with the highest drug overdose rates are characterized by higher levels of economic disadvantage and a greater proportion of service sector employment.

The white variable is defined as the percentage of those who are non-Hispanic white. This variable is consistent with CDC data and Case and Deaton's (2015) study showing drug overdose mortality to be affecting whites.

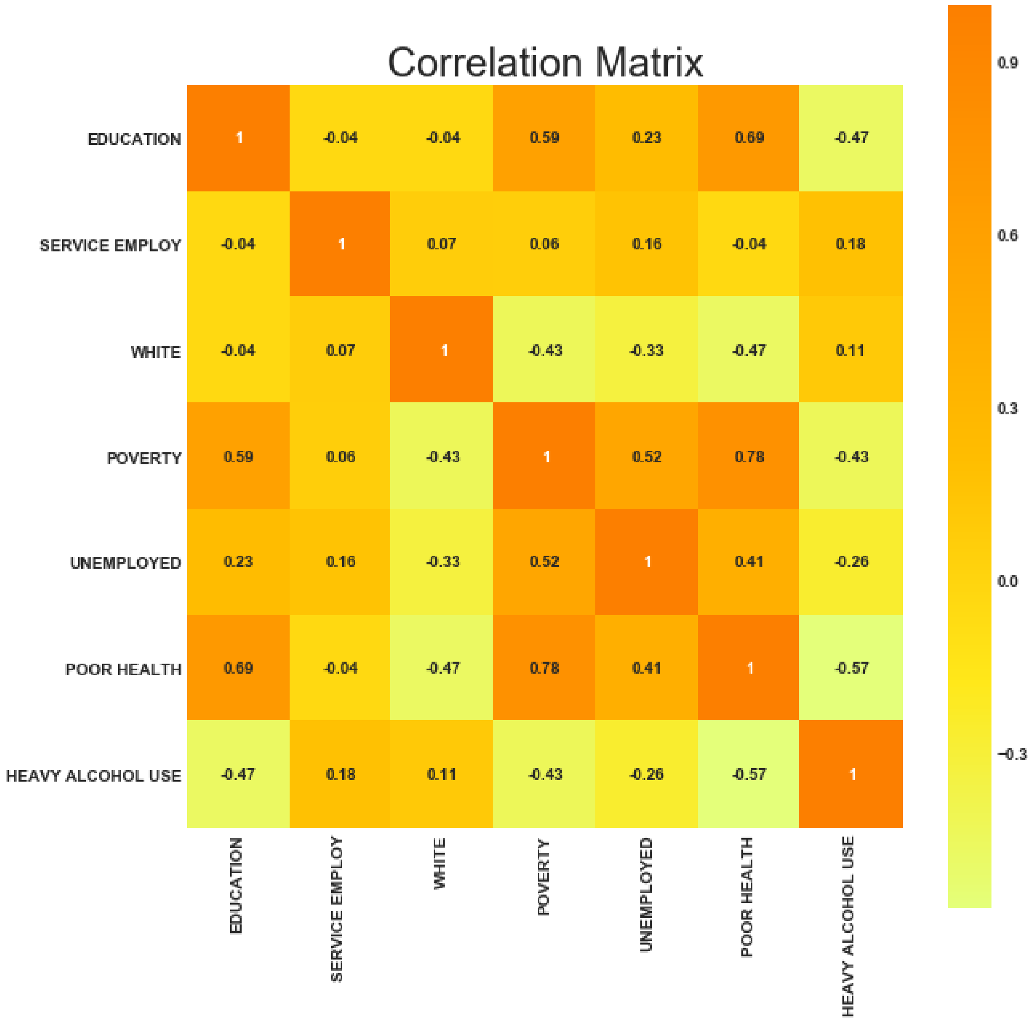
Percent reporting poor/fair health and percent reporting heavy alcohol consumption are obtained from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) and are provided at the county level. The BRFSS is a nationally representative annual telephone survey conducted by the CDC that assesses health-related risk behaviors throughout the United States. At more than 400,000 adult interviews every year, it is the largest health survey system in the world (CDC 2021).

The heavy alcohol consumption variable is defined as consuming more than 4 (women) or 5 (men) alcoholic beverages on a single occasion in the past 30 days, or heavy drinking, defined as drinking more than 1 (women) or 2 (men) drinks per day on average. For the reporting poor/fair health variable, respondents were asked “In general, would you say that your health is excellent, very good, good, fair, or poor?”. Respondents answering “poor” and “fair” were coded as poor/fair to create the poor/fair variable.

Multicollinearity

All covariates were assessed in a correlation matrix (pictured below) using Pearson’s correlation analysis to identify instances of multicollinearity. The correlation matrix reports the following: having a high school education or less is highly correlated with poor health (0.69), the poverty rate is highly correlated with poor health (0.78). Given these high correlations, all variables were assessed by their Variance Inflation Factor (VIF). The poor health variable had the highest VIF of 4.6. All other variables had VIFs below 3. A VIF score above 5 indicates moderate multicollinearity while a score above 10 indicates high multicollinearity. The results of the VIF assessment indicate that multicollinearity is not an issue in these analyses.

Figure 5. Correlation matrix of all covariates in the study



All covariates were assessed in a correlation matrix (pictured above) using Pearson's correlation analysis to identify instances of multicollinearity. The correlation matrix reports the following: having a high school education or less is highly correlated with poor health (0.69), the poverty rate is highly correlated with poor health (0.78). Given these high correlations, all variables were assessed by their Variance Inflation Factor (VIF). The poor health variable had the highest VIF of 4.6. All other variables had VIFs below 3. A VIF score above 5 indicates moderate multicollinearity while a score above 10 indicates high multicollinearity. The results of the VIF assessment indicate that multicollinearity is not an issue in these analyses.

METHODS

The data are analyzed using fixed effects linear regression with robust standard errors clustering for DMA. Two-way fixed effects control for yearly and DMA fixed effects. Model specifications are consistent with previous area-level mortality studies (Brainerd 2001, Kunce & Anderson 2002, Neumayer 2003, Vitt et al. 2018, Anestis et al. 2017, Stack & Kposowa et al. 2016, Kposowa et al. 2016, Cylus et al. 2014). Model specification was supported via Hausman test that indicated the appropriateness of using the fixed effects over the random effects model.

The multivariate fixed effects model to be estimated is as follows:

$$OR_{ij} = \beta ST_{it} + \beta X_{ij} + \gamma_t + \alpha_i + \varepsilon_{ij}$$

The term OR is the fatal drug overdose rate per 100,000 in DMA *i* at time *t.* The variable of interest, Google search terms, is represented by vector ST. Vector X consists of the following control variables: educational attainment (high school and below), workforce employed in service occupations, white, poverty, unemployment, poor health, and heavy alcohol consumption. Parameter γ represents time effects that account for unobserved national forces driving the fatal drug overdose rate. Parameter α represents unobserved DMA-specific fixed effects. The error term is represented by vector ε.

Fixed effect models control for potential omitted unobserved time-invariant area-specific factors that may affect the fatal drug overdose rate. Time-invariant factors that rarely undergo drastic changes through time are effectively controlled by the model.

These factors include culture, climate, geography, elevation, religious composition, gender, and race. DMA-level factors that may contribute to fatal drug overdose, such as a drug culture, are controlled for as well.

Time-specific social currents such as recessions or major events that can affect fatal drug overdose rates on a national level can provide misleading estimates. In efforts to not confound the estimates with an exogenous variable that may affect overdose trends, the specified model will control for such trends by controlling for yearly effects. It may be the case that high-profile overdoses may spur national increases in the overdose rate, nevertheless such instances are controlled for by the yearly fixed effects. Standard errors are clustered by DMA to address serial correlation which biases standard errors. Analyses are conducted using the fixest package in R.

RESULTS

Several models were estimated. First, bivariate linear regressions were run with each variable as the independent variable and the fatal drug overdose rate as the dependent variable. None of these models included fixed effects. Robust standard errors were clustered at the DMA level.

Bivariate models are run to demonstrate each variable's independent relationship with the fatal drug overdose rate, without controlling for other unobserved confounders. This allows us to verify each variable's association with the fatal drug overdose rate, however this leaves room for unobserved confounders. Unobserved confounders are controlled for in models after this.

The results of these bivariate models are shown in table 1 below. The drug search term composite is significant at the 0.001 level with a coefficient of 0.25 in the expected direction, and produces the largest r-squared of 0.12. The second highest r-squared is 0.02 for the service employment, white, and unemployment variables. Service employment is significant at the 0.01 level in the expected direction. Percent white is significant at the 0.05 level in the expected direction. Unemployment is significant at the 0.05 level in the expected direction.

Table 1. Bivariate linear regression. Dependent variable: fatal drug overdose rate (2010-2016)

Linear Regression (Bivariate)

Dependent Variable: Fatal Drug Overdose Rate (2010-2016)

	Coefficient	SE	R2
Drug Search Term Composite	**0.2548*** **	0.0460	0.12
Education (HS & below)	**0.0613**	0.0661	0.004
Service Employment	**1.169** **	0.3659	0.02
White	**0.0484* **	0.0197	0.02
Poverty	0.0056	0.1082	0.00
Unemployment	**0.4122* **	0.1649	0.02
Poor Health	**0.2718`**	0.1465	0.02
Alcohol	**-0.2230`**	0.1305	0.01
Fixed-Effects:	None		
SE:	Robust / Clustered by DMA		
Observations	1,463		

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Next, a multivariate linear regression was run with the fatal drug overdose rate as the dependent variable. This model does not include any fixed effects. Robust standard errors were clustered at the DMA level. The results of this multivariate model are shown in table 2 below.

This model is run to observe the relationship these covariates have with the drug overdose rate, and to observe which associations are no longer significant as a result of controlling for the covariates. It is important to note that this model does not control for other unobserved confounders. Unobserved confounders will be accounted for in the next two models.

After controlling for the most popular factors associated with overdose, the drug search term composite is significant at the 0.001 level in the expected direction. This demonstrates that the drug search term composite is significantly associated with the fatal drug overdose rate independent of the major overdose predictors from the literature. Therefore, Google search data adds to what we already know about overdose.

The education variable becomes significant at the 0.001 level, however in the unexpected direction. The service employment variable is significant at the 0.05 level in the expected direction. Percent white is significant at the 0.05 level in the expected direction. The poverty variable is significant at the 0.001 level, however in the unexpected direction. The unemployment variable is significant at the 0.001 level in the expected direction. The poor health variable is significant at the 0.001 level in the expected direction, and produces the biggest coefficient of 1.08. It must be pointed out that the poverty variable is highly correlated with the poor health variable. Their correlation coefficient is 0.78. When measuring their VIF, the poor health variable had a value of 4.6, while the poverty variable had a value of 2.3. These values indicate that multicollinearity is not an issue in these models, however the fact that poor health becomes significantly associated with the overdose rate in the multivariate model calls for further exploration of possible interaction effects, especially with poverty.

Table 2. Multivariate linear regression. Dependent variable: fatal drug overdose rate (2010-2016)

Linear Regression (Multivariate)

Dependent Variable: Fatal Drug Overdose Rate (2010-2016)

	Coefficient	SE
Drug Search Term Composite	**0.2644*** **	0.03833
Education (HS & below)	**-0.2359*** **	0.05657
Service Employment	**0.8503* **	0.30368
White	**0.0611* **	0.02390
Poverty	**-0.6689*** **	0.12624
Unemployment	**0.5928*** **	0.14178
Poor Health	**1.0823*** **	0.20931
Alcohol	**-0.15129**	0.09842
Fixed-Effects:	None	
SE:	Robust / Clustered by DMA	
Observations	1,463	
R2	0.28	
BIC	9253	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

The models above demonstrate that the data is valid and most of the variables produce expected results. The models also demonstrate that the variable of interest, the drug search term composite, is significantly associated with the fatal drug overdose rate for the period of 2010 to 2016.

The following models include yearly and DMA fixed effects. It is important to note that the fixed effects effectively control for time-invariant factors. Several of the variables in the models can be described as time-invariant, such as percent white, the education variable, reporting poor health, the unemployment rate, and heavy alcohol consumption. For example, over the course of time, the percent white of an area does not undergo drastic changes; it remains relatively stable. The fixed effects in the model will omit the effects these time-invariant factors have on the fatal drug overdose rate and as a result, the results produced by these time-invariant factors should be evaluated keeping these facts in mind. In fact, a descriptive analysis of the national-level trends of these variables shows that all of them are relatively time-invariant.

Table 3 below describes the national-level trend of each covariate for the year 2010 and 2016. Between these seven years, having a high school degree or below decreased from 44 percent to 40.5 percent. Service employment increased from 5.4 percent to 5.8 percent. The proportion of whites in the US decreased from 64.7 percent to 62 percent. The poverty rate increased from 12.6 percent to 14.2 percent. The unemployment rate decreased from 7.9 percent to 7.4 percent. Reporting poor health increased from 16.1 to 16.5 percent. Heavy alcohol consumption increased from 15 percent to 17.5 percent. From these trends, we can observe that changes in these covariates, with the exception of mental health practitioners per capita, are very likely to be captured and therefore controlled for by the fixed effects models discussed below.

Table 3. Descriptive statistics of covariates showing changes between 2010 and 2016

Descriptive Statistics

US National-Level	2010	2016
Education (HS & below)	44%	40.50%
Service Employment	5.40%	5.80%
White	64.70%	62.00%
Poverty	12.60%	14.20%
Unemployment	7.90%	7.40%
Poor Health	16.10%	16.50%
Alcohol	15%	17.50%

Source: US Census Bureau & CDC BRFSS

Bivariate linear regressions were run with each variable as the independent variable and the fatal drug overdose rate as the dependent variable. These models include both yearly and DMA fixed effects. Robust standard errors were clustered at the DMA level. These bivariate models are run to demonstrate each variable’s independent relationship with the fatal drug overdose rate, while controlling for unobserved confounders. This allows us to verify each variable's association with the overdose rate.

The results of these bivariate models are shown in table 4 below. The drug search term composite is significant at the 0.001 level with a coefficient of 0.21, and with the largest r-squared of 0.08. The r-squared of all other variables are negligible. The education variable is significant at the 0.01 level in the unexpected direction. Unemployment is significant at the 0.01 level in the unexpected direction. The alcohol consumption variable is significant at the 0.001 level in the unexpected direction. The only significant association that is in the expected direction is the drug search term composite. The other variables are significant in the unexpected direction due to the implementation of the yearly and DMA fixed effects.

Table 4. Two-way fixed effects bivariate linear regression. Dependent variable: fatal drug overdose rate (2010-2016)

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Bivariate)

Dependent Variable: Fatal Drug Overdose Rate (2010-2016)

	Coefficient	SE	R2
Drug Search Term Composite	**0.2109*****	0.03116	0.08
Education (HS & below)	**-0.8070****	0.24921	0.01
Service Employment	-0.6964	0.53400	0.003
White	0.6597	0.43730	0.005
Poverty	0.2335	0.21542	0.001
Unemployment	**-0.6907****	0.25112	0.01
Poor Health	0.2732`	0.15915	0.005
Alcohol	**-0.4077*****	0.11443	0.02
Fixed-Effects:	Year & DMA		
SE:	Robust / Clustered by DMA		
Observations	1,463		

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Lastly, a multivariate linear regression was run with the fatal drug overdose rate as the dependent variable. These models include both yearly and DMA fixed effects. Robust standard errors were clustered at the DMA level. The results of this multivariate model are shown in table 5 below. This model is run to observe the relationship the covariates have with the fatal drug overdose rate, and to see which associations are no longer significant as a result of controlling for the covariates. Additionally, this model allows us to observe which variables remain significant after controlling for unobserved confounders.

The variable of interest, the drug search term composite, is significant again at the 0.001 level producing a coefficient of 0.20. Controlling for all confounders, a ten point increase in opioid-related search interest is associated with a two point rise in the fatal drug overdose mortality rate. For an area of one million people, this equates to about 20 extra deaths, or roughly 600 years of potential life lost. Years of potential life lost is calculated by subtracting the age of suicide by the standard life expectancy of 75, then taking the conservative assumption that all deaths occurred around the age of 50 (CDC 2018).

The drug search term composite is found to be significant in the expected direction in every single model discussed so far. The two-way fixed effects produce strange results with the other significant variables including education, unemployment, and alcohol consumption. However, the model also shows that poverty is significant at the 0.05 level in the expected direction. The drug search term composite shows its reliability by remaining significant while undergoing several models while controlling for some of the most prominent predictors of overdose from the literature. Additionally, the drug search term composite, when comparing it to other variables in the bivariate results, produces a far greater r-squared than the other variables.

Table 5. Two-way fixed effects multivariate linear regression. Dependent variable: fatal drug overdose rate (2010-2016)

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Multivariate)

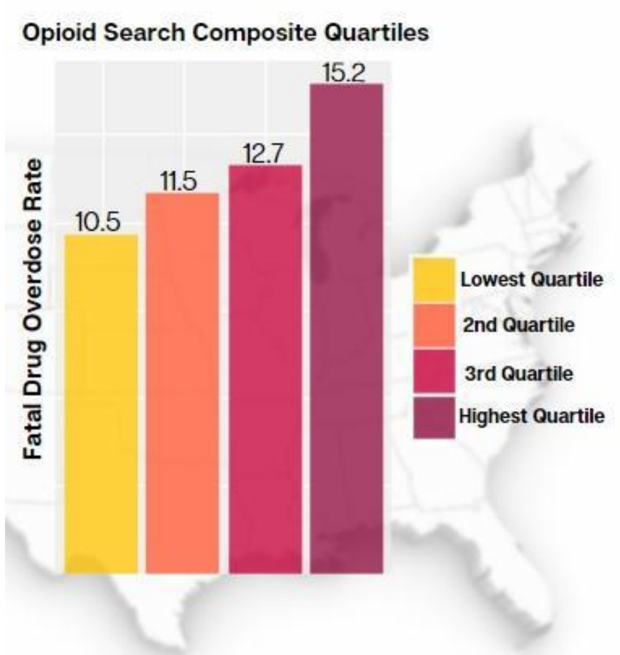
Dependent Variable: Fatal Drug Overdose Rate (2010-2016)

	Coefficient	SE
Drug Search Term Composite	**0.2013*** **	0.03018
Education (HS & below)	**-0.7341** **	0.25005
Service Employment	-0.7924	0.50121
White	0.3003	0.39004
Poverty	**0.4903***	0.19734
Unemployment	**-0.5959***	0.24272
Poor Health	0.2511`	0.14106
Alcohol	**-0.3715** **	0.11262
Fixed-Effects:	Year & DMA	
SE:	Robust / Clustered by DMA	
Observations	1,463	
R2	0.13	
BIC	8673	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

The effect of living in a DMA in the highest quartile of the drug search composite increases the risk of fatal drug overdose considerably. As pictured in figure 6 below, DMAs in the highest quartile of the drug search composite report almost a 45 percent higher rate of fatal drug overdose compared to areas in the lowest quartile. DMAs in the highest quartile of the drug search composite report an average fatal drug overdose rate of 15.2. DMAs in the lowest quartile of the drug search composite report an average fatal drug overdose rate of 10.5.

Figure 6. Graphic showing the fatal drug overdose rate through quartiles of the opioid search composite



The drug search term composite is significant in the expected direction in all models listed above. This demonstrates that it is a reliable construct that is capable of capturing beyond what is already known by the literature. Since it remains significant after controlling for the major predictors of overdose, it demonstrates that Google searches can add to what we already know about overdose. After showing that it remains significant after controlling for unobserved confounders, it shows that it is significant independently of all other factors and therefore, it can be argued that Google searches provide us with another way of observing an area's inclination towards drug overdose in and of itself.

Lagged Model

While the models discussed above assess the variables and fatal drug overdoses on a yearly basis, this approach can bring up questions of causality. Therefore, all independent variables were lagged by one year and regressed on the next year's fatal drug overdose rate. For example, the drug search term composite of 2015 was regressed on the 2016 fatal drug overdose rate. To conduct this analysis, the 2009 data for all independent variables were obtained. First, bivariate models were run with the fatal drug overdose rate as the dependent variable. These models include both yearly and DMA fixed effects. Robust standard errors were clustered at the DMA level. The results of these bivariate models are shown in the table below. The drug search term composite is significant at the 0.001 level and produces a coefficient of 0.17. The education variable is significant at the 0.01 level, however with the coefficient in the unexpected direction. No other variables are significant. The highest r-squared 0.05 is produced by the drug search term composite. All other r-squared results are negligible. This analysis gives credence to the idea that last year's drug-related Google search may have an impact on the following year's fatal drug overdose rate.

Table 6. Lagged two-way fixed effects bivariate linear regression. Dependent variable: fatal drug overdose rate (2010-2016)

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Bivariate)

Dependent Variable: Fatal Drug Overdose Rate (2010-2016)

	Coefficient	SE	R2
Drug Search Term Composite 1-yr lag	**0.1737*** **	0.029092	0.05
Education (HS & below) 1-yr lag	**-0.8485** **	0.269528	0.01
Service Employment 1-yr lag	-0.3815	0.471768	0.00
White 1-yr lag	0.3894	0.324823	0.002
Poverty 1-yr lag	0.2539	0.212116	0.002
Unemployment 1-yr lag	-0.2828	0.200162	0.002
Poor Health 1-yr lag	-0.1287	0.086798	0.002
Alcohol 1-yr lag	0.0279	0.099787	0.00
Fixed-Effects:	Year & DMA		
SE:	Robust / Clustered by DMA		
Observations	1,463		

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Next, a multivariate model is run. These models include both yearly and DMA fixed effects. Robust standard errors were clustered at the DMA level. The results of the multivariate model are shown in table 6 below. The one-year lag of the dependent variable, fatal drug overdose rate, was included in the model. The one-year lag of the drug search term composite is again significant at the 0.001 level producing a coefficient of 0.12. Not surprisingly, the one-year lag of the fatal drug overdose rate is significant at the 0.001 level. The one-year lag of the education variable is significant at the 0.01 level in the unexpected direction. The one-year lag of the poverty variable is significant at the 0.05 level in the expected direction. Again, this model, along with all models discussed above, demonstrates the robustness of the variable of interest, the drug search term composite, as it is significant in the expected direction in every model, produces the highest r-squared compared to any variable, even after controlling for the most popular predictors of fatal drug overdose. These results remain true and consistent even after lagging the predictor variables for one year, giving credence to the idea that Google search terms may have some predictive power.

Table 7. Lagged two-way fixed effects multivariate linear regression. Dependent variable: fatal drug overdose rate (2010-2016)

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Multivariate)

Dependent Variable: Fatal Drug Overdose Rate (2010-2016)

	Coefficient	SE
Drug Search Term Composite 1-yr lag	**0.1207***	0.02514
Fatal Drug Overdose Rate 1-yr lag	**0.2895***	0.05117
Education (HS & below) 1-yr lag	**-0.7832**	0.25588
Service Employment 1-yr lag	-0.1685	0.46429
White 1-yr lag	-0.0447	0.28868
Poverty 1-yr lag	**0.4530***	0.20357
Unemployment 1-yr lag	-0.4266`	0.22362
Poor Health 1-yr lag	-0.0974	0.09009
Alcohol 1-yr lag	0.0396	0.09107
Fixed-Effects:	Year & DMA	
SE:	Robust / Clustered by DMA	
Observations	1,463	
R2	0.14	
BIC	8665	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Lastly, a final model is run (Table 8). The models in Table 6 and 7 may suffer from Nickell bias (Beck et al. 2014), which occurs in panel data when there are fewer time periods than individual units. This project has 7 time periods and 209 DMA units, creating a high probability that the models in table 6 and 7 suffer from this Nickell bias. Nickell bias is caused by autocorrelation produced by lagged values which produces erroneous coefficient estimates (Arellano & Bond 1991). Thankfully the Arellano-Bond estimator overcomes Nickell bias (Arellano & Bond 1991). The lagged model was run using the Arellano-Bond estimator. The results of the model are shown below in Table 8. Once again, the model shows that the 1 year lag of the drug search term composite survives its most arduous statistical test by remaining significant at the 0.001 level while controlling for the most comprehensive predictors in the literature.

Table 8. Lagged Arellano-Bond estimator with two-way fixed effects multivariate model. Dependent variable: fatal drug overdose rate (2010-2016)

Arellano-Bond Estimator with DMA and Yearly Fixed Effects (Multivariate)

Dependent Variable: Fatal Drug Overdose Rate (2010-2016)

	Coefficient	SE
Drug Search Term Composite 1-yr lag	**0.1347*****	0.02227
Fatal Drug Overdose Rate 1-yr lag	**0.3791*****	0.03722
Education (HS & below) 1-yr lag	**-0.7074****	0.23440
Service Employment 1-yr lag	-0.0113	0.40273
White 1-yr lag	0.5275	0.30989
Poverty 1-yr lag	**0.5264***	0.20343
Unemployment 1-yr lag	**-0.4646***	0.21594
Poor Health 1-yr lag	0.0584	0.15247
Alcohol 1-yr lag	0.0064	0.10982
Fixed-Effects:	Year & DMA	
SE:	Robust / Clustered by DMA	
Observations	1,463	
R2	0.17	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Robustness Check

It is possible that the drug search composite may be confounding with an unobserved aspect of an area rather than a specific phenomenon related to drug overdose. To test this possibility, we examined other possible health outcomes, cancer mortality and heart attack mortality, as dependent variables. If the drug search composite is significantly associated with such outcomes, then this study cannot argue on the reliability of Google data and its association with the fatal drug overdose rate.

Cancer mortality and heart attack mortality rate data was obtained from the CDC compressed mortality file. Cancer mortality is defined by ICD-10 codes C00-C97. Heart attack mortality is defined by ICD-10 codes I210-I219. Cancer and heart attack mortality rates were aggregated by DMA and merged to the final dataset. The drug search composite was tested in a bivariate fixed effects model controlling for DMA and yearly fixed effects. The drug search composite did not have an association with cancer mortality (t = -0.09). The drug search composite did not have an association with heart attack mortality (t = -0.65). Therefore, these analyses strengthen the assertion that Google search term data may reveal profound social currents that dispose individuals towards succumbing to drug overdose, and that such associations are not spurious in nature.

DISCUSSION

The discussion section summarizes the findings from the result sections, focusing on the drug search term composite, and arguing that Google search data can add to what is already known from the literature. It touches on the strengths of implementing Google search composites as opposed to using single search terms. The discussion finishes by citing the importance of using Google search data and other internet sources that can provide researchers with rich information on human attitudes, behaviors, and inclinations.

This study provides substantial evidence for the utility of Google search terms to measure drug-related interest at the area level. To the best of our knowledge, this is the first study to assess the association between the frequency of drug-related searches and the fatal drug overdose rate at the DMA-level. The results of the models discussed above demonstrate that the variable of interest, the drug search term composite, is robust and reliable throughout the entire series of models, even after controlling for the most popular predictors of fatal drug overdose from the literature. Additionally, the relationship is significant over the period of seven years (2010-2016).

Because the drug search term composite remains significant after controlling for numerous covariates, it demonstrates that Google search data can reveal beyond what traditional predictors already can reveal. After controlling for unobserved confounders via the fixed effects model, the drug search composite remained significant indicating that it is significant independently of all other factors and therefore it can be argued that Google search data provides us with another way of observing an area's inclination towards overdose in and of itself. Lastly, the fact that the previous year drug search composite has a significant effect on the following year's overdose rate demonstrates that Google search data may have some predictive power.

Regarding the effect drug searches have on the rate of overdose, citing table 5, the multivariate fixed effects model, the drug search term composite produces a coefficient of 0.20. Controlling for all observed and unobserved confounders, a ten point increase in opioid-related search interest is associated with a two point rise in the fatal drug overdose mortality rate. For an area of one million people, this equates to about 20 extra deaths, or roughly 600 years of potential life lost.

The results of this analysis bring Google search terms into the spotlight as a viable predictor of fatal drug overdose. As Google search terms have been able to assess numerous health outcomes, including herpes, herpes zoster vaccinations, outbreaks of the plague, cancer, whooping cough, influenza, asthma, Lyme disease, tropical infections, syphilis, HIV, influenza hospitalizations, Zika virus, AIDS prevalence, suicide, dangerous pharmaceutical drug interactions and many more, (Mavragani & Ochoa, 2019; Arora et al. 2019, Jun et al. 2018; Brodeur et al. 2021), this study has successfully used opioid-related Google search terms to assess and predict fatal drug overdoses at the DMA level over the period of 2010 to 2016.

In summary, by using principal component analysis to create a drug search composite, and using robust fixed effects and non-fixed effects models, this study explored over 60 drug-related Google search terms to examine whether they are associated with the fatal drug overdose rate. This study identified an opioid-related composite with good internal consistency consisting of the following terms; heroin, naloxone, naltrexone, Narcan, overdose, Vicodin, vivitrol, opiate, opioid. This opioid-related composite was found to be significantly associated with the fatal drug overdose rate through the period of 2010 through 2016. This association remained significant even after both yearly and DMA-level fixed effects were implemented. This association remained significant even after controlling for the popular fatal drug overdose predictors from the literature. A significant association was observed for the one-year lag of the variable of interest, indicating that the opioid-search term composite may possess some predictive power.

Google and internet data

These findings call for further research into our online digital traces and how they may provide researchers with clues about area-level health. While causality is difficult to ascertain from these analyses, the significant associations found warrant further investigation of how area-level Google search behavior is tied to health and mortality outcomes. This study demonstrates that what areas Google about, specifically opioid-related search terms, are significantly associated with the fatal drug overdose of such areas. In other words, in areas where drug overdose is prevalent, we observe area residents Googling about these very events as they occur, before they occur, and after they occur. In a sense, observing what an area Googles about frequently gives researchers a glimpse of the social conditions, or social climate, of such areas.

These findings give credence to the idea that social climates, such as those described by Mark Hatzenbueher (social attitudes and social norms of an area), can be observed using Internet-based measures. Given that roughly 90 percent of Americans use Google as their search engine, and nearly 93 percent of Americans have internet access via computer or smartphone (Pew Research Center 2021), Google data gives researchers a rare glimpse of almost an entire area's thoughts and behaviors. The extent that Google is used by so many individuals and so frequently throughout their day, provides us an almost unparalleled source of information on social life. This study gives credence to the idea that an area's interests, thoughts, and behaviors are reflected in Google search data. These data may assist in assessing further concepts such as culture, social trends, prejudicial attitudes, and more.

In light of Google search data producing promising results throughout numerous fields, the internet provides additional sources of social information such as Twitter, Facebook, and Reddit. These sources provide rich amounts of data as they oftentimes provide detailed descriptions of individual experiences and stories. Using internet data in research has increased precipitously in the last decade, which has brought us an era of computational social science. The internet offers an unprecedented wealth of text, image, and audio data produced on numerous sites and applications, which provides us with the opportunity to catalog human history on an unprecedented scale (Mohr et al. 2020). New forms of data may allow us to identify patterns that could expand the range of research questions social scientists may ask for years to come (Mohr et al. 2020).

LIMITATIONS

Not all Google searches for drug-related content are necessarily made by people experiencing drug cravings or looking for drug use methods. Nonetheless, an area conducting a more than average search volume for drug-related content can be said to be reflecting the overall social environment of such an area. For example, Googling for "COVID symptoms" does not necessarily mean one has COVID, but it does reflect a situation, concern, or sentiment that is influenced by the surrounding social environment, and therefore constitutive of it. In the same way, an area experiencing many overdoses will have many residents either concerned, affected, or interested in the topic, thereby producing a higher-than-average search volume for drug-related searches.

While there is widespread internet usage among Americans, social scientists understand that technologies and amenities are not distributed evenly among populations. There are documented disparities in internet usage among age groups, educational attainment, income, race/ethnicity, and urbanity (Statista 2021). While internet usage ranges from 99 percent to 96 percent for the ages of 18 to 64, a mere 75 percent of adults over the age of 65 use the internet. The share of college graduates and those with some college who use the internet ranges between 98 to 97 percent however, the share of those with high school or less who use the internet is 86 percent. Given the correlation between education and income, it is no surprise that 86 percent of those who report less than 30,000 dollars yearly income are internet users, compared to 98 percent of those who report making over 50,000 dollars a year (Statista 2021). In terms of race/ethnicity, surprisingly Hispanics use the internet slightly more than whites at 95 and 93 percent respectively. 91 percent of African Americans are internet users. In terms of geographic location and environment, 95 percent of urban residents are internet users, compared to 94 percent of suburban residents, and 90 percent of rural residents. However, given these disparities, this study was still able to find a significant association between opioid-related search terms and the fatal drug overdose rate. Speculation on why this may be the case may lie in the fact that an overwhelming amount of overdose deaths occur to those ages 25 through 45, a demographic that reports high rates of internet access and usage.

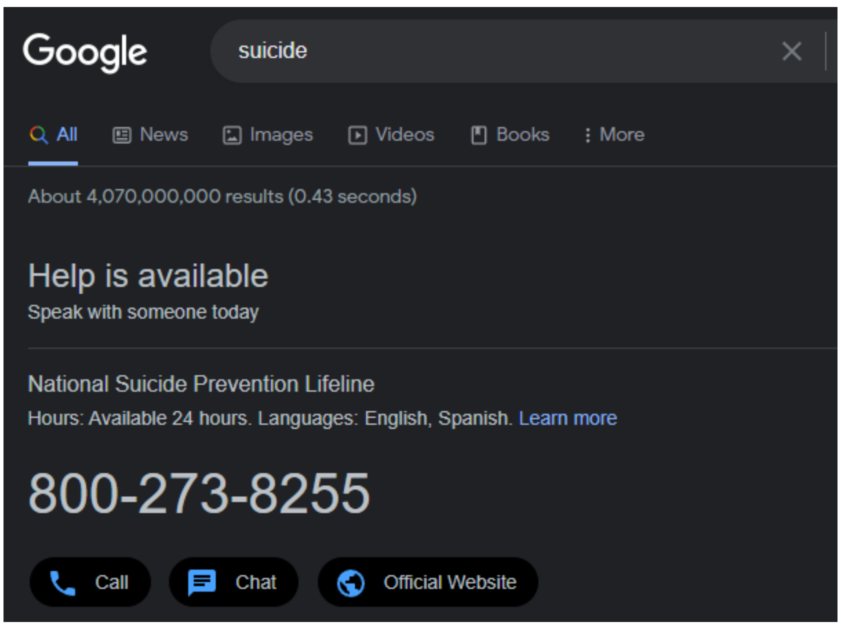
Conducting demographic group analysis, such as gender, age, and ethnicity is not possible using Google Trends. Google Trends is not able to parse out these demographic differences because of privacy concerns. Instead, the data is aggregated by area. Therefore, the search frequency score assigned to an area reflects every search conducted in that area. This methodology counts every search by every user in that area for the specified time frame, and can be thought of as an aggregate sample.

POLICY IMPLICATIONS

DMAs that frequently Google for these terms should be watched closely by the CDC. Areas that frequently Google about drug-related terms or information on drugs may indicate multiple area residents are more likely to engage in drug use and result in possible overdose. Knowledge on which areas are most at risk informs policy and health initiatives, such as where to allocate overdose prevention resources. Thankfully these data are available in real-time and therefore DMAs that experience a rise in drug-related search term frequency can be identified and targeted swiftly with prevention efforts. For example, the systematic screening of drug-related search terms, which reflects real-time search behaviors in an area, can assist in monitoring area-level drug-interest and overdose risk.

Additionally, a cost-effective and practical initiative to help ameliorate fatal drug overdoses would have health officials collaborate with Google to edit their search algorithm. This edit would involve the appearance of overdose-prevention content such as informing users where to obtain Narcan and other overdose prevention medications. A similar initiative has already been implemented when one searches for the term "suicide", with an alert appearing on the screen informing the user of suicide-related help content. This example is shown below.

Figure 7. Prompt issued by Google when searching for suicidal topics



ACADEMIC CONTRIBUTIONS

This project advances social epidemiology research by implementing an innovative methodology that is seldom used. Previous studies have validated this methodology at the state-level for drug overdoses. However, no previous studies have done these analyses at the DMA-level. This project advances sociological research in that it introduces an emerging and verified methodology to its field. This project also provides sociologists with a new way of thinking about how personal thoughts and behaviors can be assessed by Google search term data. Google can show us people's innermost thoughts, and this project demonstrates to sociologists that this data ought to be taken advantage of.

This project advances our theoretical understanding of the social field, and how effects of the social field can be measured by using Google search data. This study argues that effects produced by changes in the social field can be measured by observing what is being searched for on Google. As previously noted, changes in the social field, such as the election of Donald Trump, led to a spike in searches for "will I be deported". Similarly, changes in the social field, such as becoming unemployed, can drive individuals to consider drug use, and their sentiments are often captured in Google searches. Since this study found the drug search term composite to be robust throughout a series of models, these findings demonstrate that Google search data is capable of capturing an individual's innermost thoughts, attitudes, and inclinations. Furthermore, this source of data has the potential to complement what is already known about drug overdose risk-factors. The findings of these analyses support Google search term data as an additional lens into social behavior.

CONCLUSION

The drug-related Google search term composite is significantly associated with the DMA fatal drug overdose rate between 2010 and 2016. This analysis gives support to analyses using Google Trends as a viable source of data that captures the behaviors, attitudes, and inclinations of a population. In this instance, Google Trends is shown to better capture the variance in fatal drug overdose rates than traditional data sources and methods. The drug search composite also shows promising results for drug overdose prediction.

Association between Traditional Masculinity & Firearm-related Google Searches and the Male Suicide Rate

Abstract: Suicides have been increasing dramatically in the last two decades. Males are almost four times more likely to commit suicide than females. The US experiences one of the highest rates of male suicide in the western world. One possible hypothesis is that the culture of traditional masculinity (TM) that American society produces leaves many men who adhere to TM unable to cope with stress, crisis, and loss in their attempts to attain, maintain, and reinforce their identities. TM adherents experience prolonged identity threats, stress mis-management, emotional illiteracy, anger and violence as viable solutions to stress, and a reluctance to seek help, which forms a dangerous amalgamation that is conducive to male suicide. Another possible hypothesis is that area-level firearm interest may be associated with firearm ownership, which is associated with male suicide. Google search term data may help researchers understand variations in area-level male suicide rates by assessing the extent that areas subscribe to TM-culture and Firearm interest. This analysis tests whether TM-related and firearm-related Google searches are associated with area-level male suicide rates from 2010 through 2016 using two-way fixed effects as a means of controlling unmeasured confounding, while controlling for major male suicide predictors from the literature.

INTRODUCTION

Suicide is increasingly becoming a public health issue in the US. While several predictors from the literature provide reliable assessments of area-level mortality, these assessments are useful after suicides have occurred. This study proposes using a new lens of observing social phenomena, real-time Google search term data, to understand variations in area-level male suicide rates. Google provides information on how popular search terms are in areas over time. Therefore, **it is possible to test if Google queries for certain topics that are conducive to male suicide are associated with the male suicide rate**.

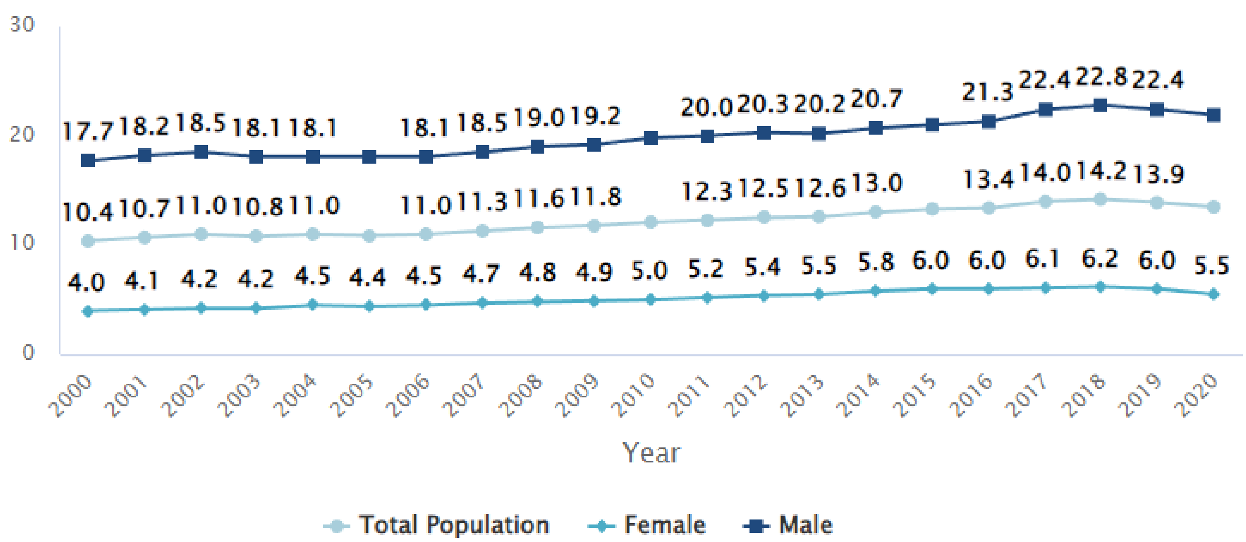
While previous studies have tested if suicidal Google searches are associated with suicide rates, this study will attempt to assess male suicide rates with less direct, and culture-related topics. Previous studies have found that the culture of traditional masculinity and firearm culture are associated with male suicide (Coleman et al. 2020, Miller et al. 2007). **Therefore, this study will test if traditional masculinity-related and firearm-related Google searches are associated with male suicides.** The level to which an area subscribes to a certain culture may allow us to assess their inclination toward suicide, especially if such culture is associated with suicide.

This paper begins with a review of the male suicide problem in the US and the major predictors from the sociological, epidemiological, psychiatric, and suicidology literature. It then describes how firearm ownership is a prevalent risk factor. It then describes a theory on why traditional masculinity is conducive to male suicide. It then describes the utility of using Google data and why it is a good and novel source of data. This section is followed by a review of previous studies that have tested the association between Google searches and suicides. That section is followed by the major contributions of this project, followed by the data, methods, results, and discussion sections.

Male Suicide is a Major Public Health Issue

The male suicide rate in the US has increased by 25 percent since the year 2000 (NIMH 2022). Suicide is the second leading cause of death for individuals between ages 10-14 and 25-34 and the third for ages 25-34 (NIMH 2022). In 2018, almost 11 million Americans seriously contemplated suicide, 3.3 million made a plan, 1.4 million attempted suicide, and nearly 50,000 completed suicide (CDC 2020). A life is lost to suicide every eleven minutes in the US (CDC 2020). Males are almost four times more likely to commit suicide than females (CDC 2020). Males are more likely to use lethal and effective means, such as firearms, while females are more likely to use less effective means such as poisoning (Schrijvers et al. 2012). Below is a chart courtesy of the National Institute of Mental Health illustrating the overall, male, and female suicide rate since the year 2000.

Figure 1. Suicide rate in the US from 2000 to 2020



Individual-level pathways to suicide include divorce (Neumayer 2003, Wray et al. 2011), unemployment (Kposowa et al. 2019, Neumayer 2003, Norstrom 1995), depression (Phillips et al. 2007), stress (Lester & Gunn 2016), and alcohol abuse (Neumayer 2003, Norstrom 1995, Kaplan et al. 2016). However, these predictors have produced mixed results in recent years (Wray et al. 2011). Having a history of previous attempts is one the most prominent and consistent predictors of suicide, however this data is difficult to procure (Owens et al. 2002).

More consistent area-level pathways to suicide include firearm ownership, having a firearm in the household, and storing firearms unlocked (Knopov et al. 2018, Kposowa et al. 2016, Vitt et al. 2018). Given that over half of suicides in the US are conducted via firearm, firearm ownership data provides a consistent estimate of suicides and firearm suicides in the US. Additional risk factors include psychiatric, substance dependency, and substance use disorders (Miller et al. 2012)

Mental illness, specifically depression, is associated with suicide (Hawton et al. 2013). While it has been found that around 90 percent of individuals who commit suicide had a mental disorder (Bertolote & Fleischmann 2002), it is also the case that the vast majority (98%) of individuals who experience mental disorders do not die by suicide (Nordentoft et al. 2011). A systematic review of suicide literature found that depression is the most common psychiatric disorder for those who die by suicide (Hawton et al. 2013). However, men are less likely to be diagnosed with or report depression compared to females (Borowsky et al. 2000).

Excessive alcohol consumption is associated with suicide (Kaplan et al. 2016, Neumayer 2003). A meta-analysis found that acute alcohol use (consuming excessive amounts prior to a behavior) is a major risk factor for attempting suicide (Borges et al. 2017). Kaplan et al. (2013) found that almost 25 percent of men who had completed suicide were found to have alcohol blood concentration. Alcohol intoxication may facilitate completed suicide by disinhibiting or diminishing an individual’s aversion and fearful response to self-harm. Additionally, chronic and heavy alcohol consumption may be an indicator of underlying mental illness, which is tied to suicide (Miller et al. 2012).

Chronic pain, regardless of type, is a risk factor for suicidality (Racine 2018, Hassett et al. 2014). Physical health conditions are found to be associated with higher odds for reporting suicide ideation, a suicide plan, and suicide attempts (Stickley et al. 2020). It is hypothesized that chronic pain may facilitate the development of depression, hopelessness, increase the desire of escape by death, and erode the fear of dying (Hooley et al. 2014). Chronic pain may create severely negative physiological and psychological conditions that an individual may be incentivized to escape from via suicide. Chronic pain may also increase an individual's tolerance of pain, making it easier to endure painful processes while attempting suicide.

Being of veteran status is a risk factor for suicide (Cerel et al. 2015, McCarten et al. 2015). It is hypothesized that military service may decrease an individual's perceptions of danger, self-harm, and fear of death, thereby increasing one's chances of attempting suicide (Joiner 2005). Additionally, combat exposure is hypothesized to contribute to fearlessness and habituation to death (Wolfe-Clark & Bryan 2017). A recent prominent empirical finding is the prevalence of physical health conditions among veterans who have committed suicide (Wood et al. 2020). This study indicates that military service has a high rate of physical injuries while military culture values physical strength and capacity. Being injured while participating in a culture that values physical strength may produce a sense of "perceived burdensomeness" that may be conducive to suicide (Joiner et al. 2009). Additionally, veterans experience high rates of posttraumatic stress disorder and major depressive disorder, which is tied suicide (Nichter et al. 2019)

The classical sociological risk factors for suicide which includes divorce, unemployment, and marriage produce mixed associations with suicide (Norstrom 1995). Using the National Longitudinal Mortality Survey, Kposowa (2000) found that from 1979 to 1989, higher rates of suicide were found in the divorced. Using the same dataset, Kposowa et al. (2020) found that from 1990 to 2011, the divorced and separated are 88 percent more likely to complete suicide than their married counterparts. Recently, unemployment was found to be significantly associated with suicide (Kposowa et al. 2019). However, these factors produce mixed results as consistent findings are elusive throughout the literature.

Divorce and unemployment are described as anomic events that reduce an individual’s social integration in society by reducing the strength of their social bonds, therefore increasing the likelihood of suicide (Durkheim 1951[1897]). The loss of a relationship, or losing a job, decreases an individual’s social integration by diminishing the strength and number of social bonds they have with others. These events further prevent society from regulating an individual’s behaviors and desires, which can lead to normlessness. For example, a job provides an individual with a structure, rules, schedule, and a sense of accomplishment. When one loses their job, they may likely lose such sources of regulation, which makes it more difficult for society to regulate their behaviors.

Marriage is generally seen as a protective factor as it further integrates one into society and forms stronger social bonds. Essentially, the greater the intensity of social bonds one possesses in society (marriage, childbearing, employment, church participation, etc.) the less likely one is to suicide. For example, being married and having children depend on an individual makes it less likely for that individual to choose to complete suicide, especially if their family depends on them. Those who lack these bonds (single, live alone, unemployed, etc.) signal a greater potential for feelings of loneliness and lack of social integration, thereby increasing the risk of suicide (Neumayer 2003). These individuals may be more likely to complete suicide as their lack a partner or children that depend on them. Essentially, it is easier to exit existence knowing fewer people will be affected. An understudied variable that may contribute to lower social integration is living alone, which this study includes. For many decades the number of Americans living in single households has increased.

Major depressive disorder, substance use disorders, and serious psychological distress produce modest associations with suicide (Miller et al. 2012). When correlated to the 50-state suicide rate, major depressive disorder, substance use disorders, and serious psychological distress produce a coefficient of 0.4, 0.2, and 0.2 respectively (Miller et al 2017). While it has been found that around 90 percent of individuals who commit suicide had a mental disorder (Bertolote & Fleischmann 2002), it is also the case that the vast majority (98%) of individuals who experience mental disorders do not die by suicide (Nordentoft et al. 2011). A systematic review of the suicide literature found that depression is the most common psychiatric disorder for those who die by suicide (Hawton et al. 2013). Not having mental health resources to treat such mental illness would further increase one's likelihood of completing suicide.

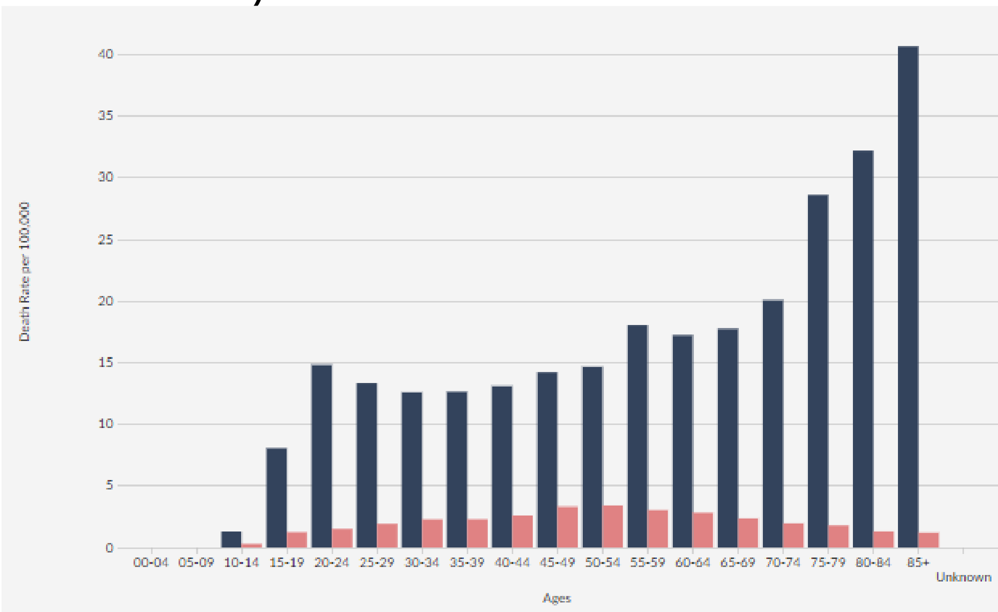
Having a history of suicide attempts has some of the strongest associations with suicide (Owens et al. 2002), however in the same respect as mental illness, its predictive ability is diminished by the fact that most people with a history of suicide attempts do not go on to commit suicide. Furthermore, while both mental illness and suicide attempt history work best as predictors at the individual level, limited data exists on these individuals, and even fewer data exists on these metrics at aggregate levels.

Firearms and Suicide

In the US, firearms are consistently responsible for 50 percent of all suicides. However for males, firearms are responsible for almost 60 percent (NIMH 2022). Suffocation, which includes hanging, is responsible for roughly 27 percent of all male suicides. Intentional drug poisoning and overdose is responsible for roughly 8 percent of male suicides. Falling or jumping from a deadly height and other forms of self-inflicted injury such as cutting or stabbing is responsible for about 7 percent of male suicides. Therefore, firearms play a major role in facilitating suicide in the US (CDC WISQARS 2020).

The male firearm suicide rate is 12.5, while the female suicide rate is 2, indicating that males are about six times more likely to use a firearm for suicide (CDC WISQARS 2020). There exists a steady age gradient in the suicide by firearm rate for males, with the exception of a blip for males ages 20-24 and 55-59. There is a precipitous rise in suicide by firearm in males after the age of 70. There is a steady age gradient in the suicide by firearm rate for females that peaks at the age of 50-54, and then steadily declines thereafter (CDC WISQARS 2020). The following is visualized below with males shown in blue and females shown in red. While this study acknowledges age differences, it will analyze the male suicide rate of all ages.

Figure 2. Differences in male and female suicide rates by age category (suicide rates from 2010 to 2020)



The US has the highest rate of suicide by firearm of any developed country (Briggs & Tabarrok 2014). It is difficult to estimate the actual number of firearms and how many households in the US own firearms as firearm registration is not required in most US states. Most recent surveys suggest there are about 393 million firearms in the US (American Gun Facts 2021). These surveys also suggest that roughly 40 percent of US households have at least one firearm in the household. According to these statistics, the US ranks number one by far in firearms per capita. While the US has just 4 percent of the world's population, it owns about 40 percent of civilian-owned firearms worldwide. Middle-aged individuals, males, and whites are more likely to own firearms (Kleck 2018). Individuals who reside in areas of smaller population or rural areas are more likely to own firearms. Individuals who engage in alcohol abuse or heavy drinking are more likely to own firearms. Individuals without college degrees are more likely to own firearms (Kleck 2018).

In the US today, most firearm deaths are suicides, not homicides (Kleck 2018). An overwhelming number of studies find an association between firearm prevalence and the firearm suicide rate (Miller et al. 2012, Knopov et al. 2019, Kposowa et al. 2016, Vitt et al. 2018, Briggs & Tabarrok 2014). Furthermore, there is an overwhelming amount of evidence indicating that having a firearm in the home is associated with a 2 to 10 times risk increase of completing suicide, depending on gender and age group (Miller et al. 2012, Mann & Michel 2016, Wiebe 2003). Regarding households that own firearms, several studies find that the risk of suicide increases when firearms are stored unsafely, left loaded, and easily accessible (Anestis et al. 2017). These findings suggest that suicide by firearm is largely an impulsive act that relies on the immediate access of a lethal means (Kposowa et al. 2016). However, there is no theory nor data to contend that firearm ownership is associated with suicidal ideation. Several studies have demonstrated that there is no association between firearm ownership and suicide ideation (Anestis & Houtsma 2018). Simply put, firearms do more make individuals more or less suicidal, firearms just provide an easily accessible and effective means of suicide.

Therefore, one of the major reasons why firearm prevalence is associated with the suicide rate is because of the high lethality rate firearms have. Males are consistently 2 times more likely to use a firearm compared to females when committing suicide (NIMH 2022). Firearms are used in about 60 percent of male suicides, compared to being used in 30 percent of female suicides (NIMH 2022). Firearms provide a largely efficient method of suicide with about a 95 percent effectivity rate (Kleck 2018). The firearm effectiveness rate is tremendous when considering the effectiveness rate for drowning is 67 percent, 78 percent for hanging, and a mere 2 percent for intentional overdoses (Mann & Michel 2016). Therefore, access to an effective lethal mean will increase the odds of a successful suicide.

There is increasing debate about how impulsive suicide is. Firearms require very little reparation and time to commit suicide. Therefore, firearms are not only extremely lethal, but also effective in terms of shortening the time to complete the suicide. Whereas other methods, such as hanging and poisoning, require some preparation, planning, and time to execute. It is very well possible that one's desire to die may not persist long enough to carry out the suicide, especially when using a method that requires a longer preparation time (Kleck 2018). Deisenhammer et al. (2009) studied a sample of suicide attempt survivors and found that 74 percent of them reported taking less than 10 minutes from when they decided to commit suicide to the time they made their suicide attempt. About 89 percent took no more than an hour to attempt their suicide. It is very well possible that hanging or jumping methods are feasible in under an hour. Therefore, it appears that suicide attempts are bound to occur, however the lethality of the method determines if the suicide is successful or not.

This being so, it is logical to think that areas that display high levels of interest in firearms are likely to have a greater prevalence of firearm owners. And because of the association between firearm ownership and suicide, these areas are likely to have higher rates of suicide. This research will analyze aggregate-level firearm interest and its association with the male suicide rate. The analysis will assess if an area with high levels of firearm interest, via Google searches, has an influence on the suicidal behaviors of the area's male residents.

Traditional Masculinity and Suicide

There may be further underlying causes to the steep rise in suicides that has occurred in the last two decades. One possible hypothesis is that the culture of masculinity that American society produces leaves many men who adhere to this masculinity unable to cope with stress, crisis, and loss in their attempts to attain, maintain, and reinforce their hypermasculine identities. Traditional masculinity (TM) is a form of masculinity endemic to American culture that aligns with conservative values, beliefs, and norms. TM provides a set of social norms that values competition, strength, emotion avoidance, femininity avoidance, and the acceptability of anger and violence (Coleman 2015).

The negative effects of TM are an absence of empathy and emotions other than anger, along with misogynistic beliefs, homophobic beliefs, violence, and an emphasis on competition and dominance (Coleman et al. 2011). TM, therefore is aversive to the individual rights of women, exemplified by holding abortion-restrictive beliefs, and aversive to the rights of LGBT individuals, exemplified by holding beliefs that restrict freedoms for LGBT individuals. Therefore, individuals who strongly adhere to TM are more likely to use misogynistic, homophobic, and hateful speech, and immerse themselves in masculine-affirming symbols such as firearms, weaponry, large trucks, and masculine-affirming activities such as hunting, fishing, and off-roading.

Men who adhere to TM may be more likely to complete suicide. Males whose identities adhere to TM may find themselves under additional stress given the rapid liberalization of global values and beliefs. Possessing a reactionary identity during a time when global attitudes and values are becoming increasingly progressive may produce a “cultural marginalization” in reactionaries, thereby increasing their levels of societal alienation and increasing the frequency of identity threats (Kalish & Kimmel 2010). Through a Durkheimian perspective, these men are experiencing a breakdown and replacement of their societal norms (anomie) and are becoming further disintegrated from society (egoism). Holding reactionary beliefs in an ever-changing world produces numerous strains, especially if these changes directly conflict with one's identity. TM stands in stark opposition to progressive attitudes, which further allows men to reaffirm and double-down on their TM identities, which further exacerbates their mental state.

TM encourages the use of violence and anger as a solution to a challenge on one's identity (Coleman 2015). Frequent and prolonged identity threats may increase the likelihood of acting out violently. Some of the common pathways to which the TM identity is challenged includes failing to live up to the masculine identity (Canetto 1997, Girard 1993), the increasing number of women in the workforce (Cohany & Sok 2007), the increasing number of women in breadwinning positions (Fry 2010), and the increasing number of individuals who identify as LGBT. Threats to one's masculinity and loss of status are shown to increase the body's physiological stress response (Taylor 2014, Sapolsky 2005). Furthermore, research shows that males who lose their breadwinner status in marriage experience poorer overall physical health (Springer et al. 2019)

Men who strongly adhere to TM experience greater amounts of psychological distress and depression (Hayes & Mahalik 2000, Zamarripa et al. 2003). Given the stress and psychological distress experienced from identity threats, TM creates a culturally conditioned narrowing of perceived options and cognitive rigidity when under stress that increases the risk of male suicide (Coleman et al. 2011). TM effectively limits males' options when faced with stress, crisis or loss, thereby increasing the likelihood of performing self-harming behaviors. As TM limits the availability of solutions to stress, oftentimes the only solution available to TM adherents is violence, or in this case, violence turned toward the self. The likelihood of thwarting self-harming behavior is greatly reduced, as TM is associated with a reluctance to seek help (Vogel et al. 2011).

The prolonged identity threats, stress mis-management, emotional illiteracy (Gregory 2012), anger and violence as viable solutions to stress, and reluctance to seek help forms a dangerous amalgamation that this article argues, is conducive to male suicide. Suicide attempters are more likely to adhere to TM and traditional gender roles (Houle et al. 2008), however the majority of these studies are done at the individual-level. Very few studies have assessed TM at an aggregate level (Alden & Parker 2005, Kaasa et al. 2014).

This research will analyze area-level TM and its association with the male suicide rate. Area-level TM adherence will be assessed via Google search terms that are characteristic of TM culture. The analysis will assess whether an area with high levels of TM influence the suicidal behaviors of the area's male residents. There is a dearth of studies that investigate this phenomenon at an aggregate or "cultural" level.

Measuring Area-level Traditional Masculinity Using Google Trends

Many studies, especially in the psychology literature, measure TM using surveys, questionnaires, and interviews. However, the majority of these studies are conducted at the individual level. Very few studies have attempted to measure TM at an aggregate-level. A notable study by Alden & Parker (2005) measured gender-role ideology and homophobic attitudes at the city-level using the General Social Survey (GSS). This study found that homophobic attitudes, specifically from a morality standpoint, are significant predictors of hate crimes. Another study observed regional cultural differences in masculinity among several European countries (Kaasa et al. 2014). Masculinity was defined as positive, orientation to achievement and competition, subordination of women, stricter doctrines, and intolerance. Accurately measuring TM over time at aggregate levels presents a methodological challenge. However, some clues regarding TM and its attitudes may lie in what people search for when they utilize Google.

Leveraging the Power of Google

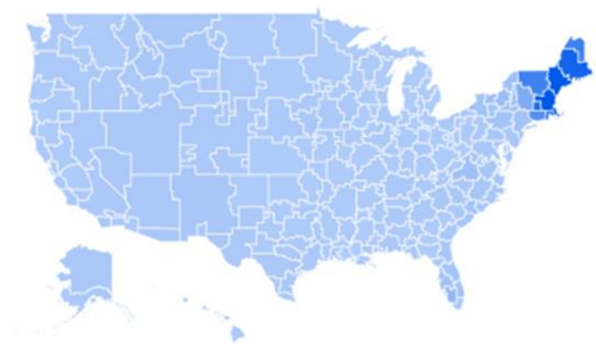
Considering the substantial and sustained increases in rates of suicide that require a deeper understanding to guide policy and to inform interventions, this research seeks to leverage a potential lens into the behaviors, activities, and sentiments of individuals. This work rests on the fact that people will ask Google almost anything. In the privacy of their searching, they will, for example, ask whether their husband is gay, if their partner is cheating on them, where confederate flags can be procured, or for anti-Black jokes. We can expect people to ask Google about things that they would not ask their parents, their partner, or their best friend (Stevens-Davidowitz 2017). It follows that Google search data may provide a useful lens into matters that would otherwise be hidden from the epidemiological and sociological inquiry. The current research seeks to assess the potential of Google search data for understanding area-level suicide mortality, through TM-related and firearm-related searches.

What makes Google a very attractive data source is the fact that Google has over 90 percent of the market share in online searches in the US, with over 92 percent of Americans being active internet users (Statista 2021). Google is the world's most utilized search engine, handling roughly 2 trillion searches per year, 167 billion searches per month, and 5.5 billion searches per day (Jun et al. 2018), making it the world's largest data source available. Given these usage statistics, Google search data has provided promising results and has grown rapidly in its use. There has, in fact, been a 20-fold increase in research articles using Google search data from 2009 to 2018 (Arora et al. 2019).

So far, research has shown novel use of Google search data including monitoring the unemployment rate (D'Amuri & Marcucci 2017), predicting the inflow of tourism, tracking home-buying interest, predicting car sales (Choi & Varian 2012, Arora et al. 2019, Brodeur et al. 2021), and predicting the direction of the Dow Jones and the S&P 500 (Hu et al. 2018). A more dismal side of this research emerges when studies assess area-level prejudice. Areas that searched for racially-charged search terms more frequently were also found to be some of the worst-performing districts for Barack Obama during the 2012 presidential elections (Stephens- Davidowitz 2014). Additionally, areas that Googled the racist n-word were associated with an 8.2 percent increase in the all-cause Black mortality rate in that area (Chae et al. 2015). These startling findings show us how aggregate Google searches have the potential to reveal meaningful social patterns and phenomena across geographic areas.

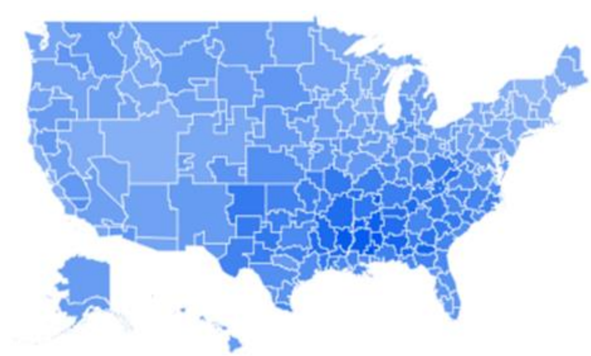
Therefore, if area-level Google searches for racially-charged terms are associated with negative political and health outcomes for African-Americans, then it is indicative that area-level Google searches provide a window into the overall culture of such areas. A more benign example below illustrates the area-level popularity of the term “Red Sox” with higher search frequencies in darker blue.

Figure 3. Geographic distribution of Google searches for “Red Sox” (last 5 years)



Since these areas in darker blue Google “Red Sox” at a far greater frequency than anywhere else in the nation, it is indicative of the sports culture of these areas, and the general interest these areas have in the Red Sox. Furthermore, we are far more likely to find Red Sox fans in these areas. In another example, if we observe the popularity of the term “bible” pictured below, we can observe a cultural trend where the most religious Americans are concentrated in the south.

Figure 4. Geographic distribution of Google searches for “bible” (last 5 years)



If these cultural trends can be observed by Google, **then this project argues that Traditional Masculinity cultural trends can also be observed, along with firearm interest, through terms that are characteristic of it.** Being that the culture of Traditional Masculinity is conducive to suicide, areas that conduct more TM-related searches should have higher rates of male suicide. Put another way, areas that subscribe to TM are more likely to search TM-related topics, and these areas should have higher rates of male suicide. Similarly, given the association between firearm ownership and suicide, areas that conduct more firearm-related searches should have higher rates of firearm ownership, which is associated with male suicide. Put another way, areas that have a higher prevalence of firearm owners are likely to search for firearm-related content, and therefore these areas should have higher male rates of suicide.

While several studies have successfully observed the association between suicide-related Google searches and suicide in Japan, South Korea, Taiwan, Germany, US, among others (Hagihara et al. 2011, Sueki 2012, Song et al. 2014, Chang et al. 2015, Paul et al. 2017, Gunn & Lester 2013, Parker et al. 2017, Capron et al. 2021, Lee 2020, Adam-Troian & Arciszewski 202, Guerra 2019), **none have examined the association of cultures to suicide rates**. Searches for “commit suicide”, ”how to suicide”, “suicide prevention”, “suicide hotline”, “suicidal”, and more, are associated with the national-level and state-level suicide rate (Gunn & Lester 2013, Guerra 2019, Capron et al. 2021). Regarding firearm-related Google searches, there have been only two studies that have observed this association. Capron et al. (2021) found that firearm suicide-related Google searches are associated with the national-level suicide rate from 2004 through 2016. These terms included: ‘how to buy a gun’, ‘how to buy a gun’, ‘where to buy a gun', 'where to buy a gun', and 'where to buy guns online'. Lee et al. (2021) found the term "gun suicide" to be correlated with the national-level firearm suicide rate from 2004 to 2017.

While the terms discussed above allow us to assess terms with very direct intention and meaning, it allows for the discussion about if less direct search terms have an association with suicide. This analysis attempts to assess a cultural and interest-related association between the frequency of Google searches and male suicide. Adherence to TM culture may be revealed by the frequency of Google search terms that are misogynistic, homophobic, intolerant, xenophobic, and associated with reactionary interests. Additionally, the extent to which an area demonstrates firearm interest should be indicative of the prevalence of firearm ownership in such areas. As firearm ownership is associated with male suicide, this study would test if area-level firearm interest, shown through Google searches, is associated with male suicide rates.

Therefore, **areas that frequently Google for TM-related terms are inferred to possess higher levels of TM culture, which this study argues is associated with higher rates of male suicide. Additionally, areas that frequently Google for firearm-related terms are inferred to have high firearm ownership rates, which this study argues is associated with higher rates of male suicide.**

HYPOTHESES

H1: Area-level traditional masculinity is positively associated with the male suicide rate independent of known area-level risk factors.

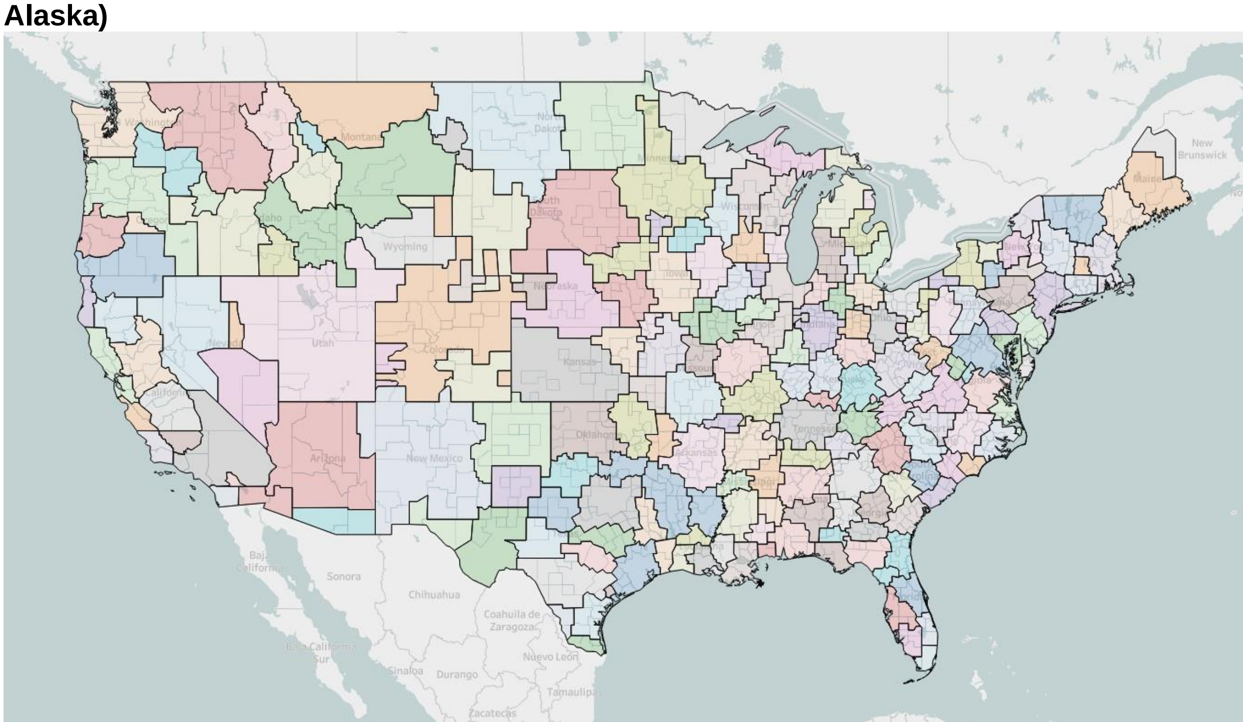
H2: Area-level firearm interest is positively associated with the male suicide rate independent of known area-level risk factors.

DATA

Unit of Analysis

This analysis is done at the Designated Market Area (DMA). There are 210 DMAs in the US. DMAs are designed for media marketing purposes and thus individuals within these DMAs are expected to share similar characteristics, interests, and profiles. The DMA is the finest geographic level currently provided by Google. One DMA (Fairbanks AK) was dropped from the analysis due to large amounts of missing data. Missing data occurs as a result of low search volume in an area. A DMA map is shown below provided by MapBox. Alaska and Hawaii are included in the DMAs however they are not pictured in this illustration. DMAs are given different colors to make viewing easier.

Figure 5. Map of Designated Market Areas in the US (Not pictured: Hawaii & Alaska)



Time

This longitudinal analysis is from 2010 to 2016. The year 2010 is chosen because it reflects the widespread adoption of smartphones, which significantly increased the frequency of Google searches. The analysis ends in 2016 as a result of the CDC Compressed Mortality File ending in 2016.

Number of Observations

As there are 209 DMAs and 7 years of data, there are 1463 unique observations.

Independent Variable

Traditional masculinity-related and firearm-related search term frequency is obtained from Google Trends. Google Trends is a data tool that provides real-time and past data regarding Google searches. Google Trends provides the frequency that a term has been searched, thereby gauging the popularity of such a term over time and in an area. Google Trends data are not user identifiable and are used here at the aggregate DMA level.

Google Trends indicates how frequently a term is searched relative to the total of all other terms searched in the area. This method ensures that areas with the highest populations will not be given the highest scores. The DMA that has the highest frequency of searches relative to the overall total number of searches is assigned a value of 100. The rest of the DMAs are assigned a value of search term frequency that is relative to the DMA with the highest value. This method is consistent with several other studies using Google Trends (Mavragani et al. 2018, Lee 2020, Prado-Roman et al. 2021, Hall et al. 2020).

Traditional Masculinity-related and Firearm-related Google search terms are guided by probabilistic, theoretical, and empirical postulations of key words an individual that subscribes to TM or firearm culture is likely to search for. While TM culture has not been previously assessed via Google search data, this project uses all firearm-related search terms from previous studies. This study recognizes that there is no standard method that produces the potential search terms, and acknowledges that other studies have used similar brainstorming procedures. Additionally, this study recognizes that a standardized method for choosing which search terms to obtain would benefit the entire field that uses Google search data.

Terms that had no more than 35 percent of its data missing were considered for the analysis. Missing Google Trends score data was imputed using Kalman smoothing. Kalman smoothing is a popular and powerful algorithm that identifies patterns in time series data and creates values based on these patterns (Li et al. 2015).

The TM-related search terms that met the missing data criteria include the following:

•aggression	•anger	•angry	•bitch
•black jokes	•breitbart	•civil war	•confederacy
•confederate	•conservative	•cunt	•don’t tread on me
•dyke	•faggot	•fag	•fox news
•foxnews	•free speech	•gadsden flag	•gay jokes
•hateful	•hate	•hitler	•homo

•homosexual	•housewife	•illegals	•info wars
•infowars	•kkk	•libertarian	•nigger
•porn	•pussy	•racism	•racist
•racist jokes	•republican	•reverse racism	•sissy
•slut	•twat	•violence	•violent
•white genocide	•white power	•white pride	•white privilege
•whore			

The Firearm-related search terms that met the missing data criteria include the following:

•ak47	•ak 47	•ammo	•ammunition
•ar15	•ar 15	•beretta	•browning
•bullets	•bullet	•bushmaster	•colt
•firearm	•firearms	•FN	•glock
•gun control	•gun shop	•gun show	•gun
•guns	•handgun	•holster	•mauser
•mossberg	•NRA	•pistol	•recoil
•remington	•revolver	•rifle	•ruger

•S&W	•shooting range	•shooting	•shotgun
•shotguns	•sig sauer	•taurus	•trigger
•weapon	•9mm		

Because these analyses use an exhaustive amount of search terms, using all search terms in the analyses increases the likelihood of model overfitting, which hinders the model's performance when conducting predictive analyses. In order to prevent model overfitting, and for the sake of model simplicity, this analysis implements a principal component analysis of the search terms of interest. Principal component analysis is a dimensionality-reduction method that reduces the amount of variables in a dataset into a lesser number, while preserving most of the information contained in the data (Abdi & Williams 2010). Principal component analysis identifies variability among correlated variables to produce a small number of variables called components. All search terms of interest were analyzed via principal component analysis for each year from 2010 to 2016. The terms that produced a loading higher than .70 on the first component through at least five years of the analysis were selected to be a part of the search term composite.

Using an exhaustive set of search terms is an improvement from previous search term studies because those studies analyzed DMA-level associations with single terms. For example, Chan (2019) used the search term “Hitler” and ‘kkk” to assess DMA-level intolerance, Hall et al. (2020) used the search term “pollen” to assess seasonal patterns in DMA-area level pollen concentration, Cousins et al. (2020) used a few COVID-19 - related search terms such as “am I sick”, “covid symptoms”, and “covid testing” to assess DMA-level COVID-19 cases, and Chae et al. (2018) used the n-word search term to assess DMA-level racist social climates.

PCA Results

Principal component analysis was conducted for every year of the analysis (2010-2016) using the search terms listed above. The analyses found that the following terms loaded on the first component with at least a .70 score for at least five of the seven years. The terms within the TM component are the following: **fox news, illegals, libertarian, foxnews, white power, whore, civil war, pussy, confederacy, don't tread on me, bitch, racism, infowars, info wars, twat, slut, cunt, and gadsden flag.** All 18 search terms in the composite produce an alpha score of 0.75, which indicates good reliability. The terms within the firearm component are the following: **bullets, ammo, S&W, ruger, holster, colt, 9mm, shotgun, gun, glock, and beretta.** All 11 search terms in the composite produce an alpha score of 0.94, which indicates excellent reliability. Alpha scores below 0.60 are considered unreliable (Vaske et al. 2017). The TM composite was created by summing the TM term scores then dividing by 18. The firearm composite was created by summing the firearm term scores then dividing by 11.

Dependent Variable

The dependent variable is the DMA-level male suicide rate per 100,000 during the years 2010 through 2016. Suicide is defined as intentional self-harm and defined by ICD-10 cause codes X60-X84. Suicide rate data was obtained from the Center for Disease Control's Compressed Mortality File. The male suicide rate was then aggregated by each DMA.

The Compressed Mortality File provides suicide counts by county. Counties are merged to their respective DMAs using a DMA index number. Each DMA has several counties within them. County male suicide counts are then aggregated by their respective DMA to produce a sum of male suicides by DMA. County total male population is also aggregated by the respective DMA to produce a DMA total male population. DMA male suicides are then divided by the total male DMA population. This number is then multiplied by 100,000 to produce DMA-level male suicide rates per 100,000 persons. County male population denominator counts are obtained from the CDC's Population File.

A very small number of male suicides in DMAs with very low populations produced extremely high suicide rates which are considered outliers. Outliers that are over 3.5 standard deviations are truncated and given values at the 3.5 standard deviation mark. This is done so that outliers do not significantly affect the regression coefficient results.

Covariates

In order to test if TM and firearm-related Google searches reveal more than what prominent male suicide predictors already reveal, this project implements a maximally robust set of control variables. This project controls for the following covariates:

•Percent married
•Percent divorced
•Percent male unemployed
•Percent veteran
•Percent living alone
•Percent without health insurance
•Percent reporting poor/fair health
•Percent reporting heavy alcohol consumption
•Primary care physicians per 100,000 population
•Mental health providers per 100,000 population
•DMA fixed effects
•Yearly fixed effects

Since all control variables are obtained at the county level, they are aggregated to the DMA-level using county population as weights. This was done by multiplying the county variable with its respective county population. This product was then summed with the other products from the same DMA. This sum was then divided by the total population of the DMA to produce an average DMA value that takes into account counties with differing populations. This procedure can also be done by taking the county population and dividing it by the DMA population. The quotient is then multiplied by the respective county variable. This procedure is repeated for each county in a given DMA. Then these products are summed up to produce an average DMA value that takes into account counties with differing populations. This procedure and the one mentioned prior were conducted and achieved identical results.

Percent married, percent divorced, percent males unemployed, percent veteran, percent living alone, and percent without health insurance are obtained from the US Census American Communities Survey 5-year estimates and are provided at the county level.

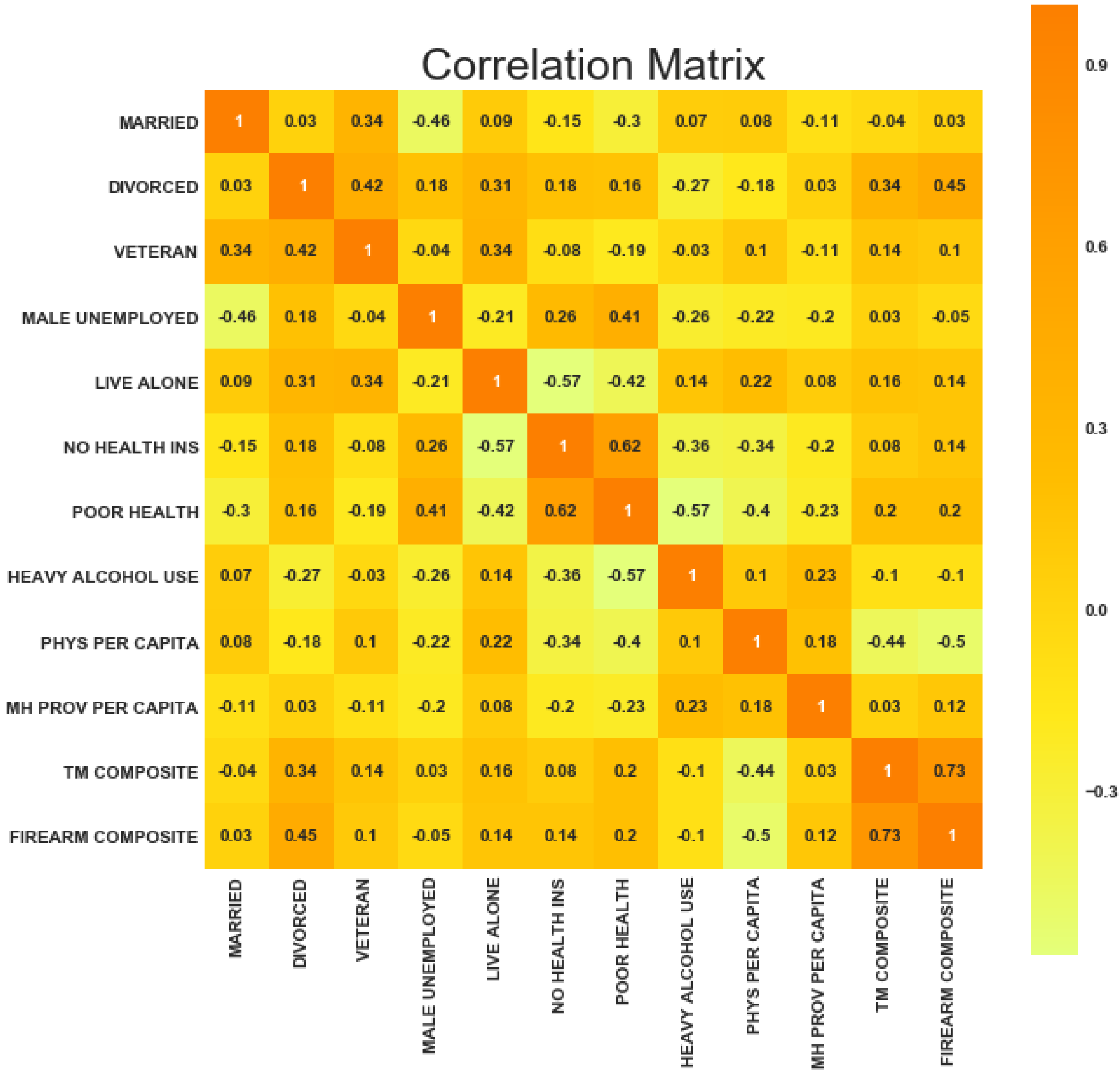
Percent reporting poor/fair health, percent reporting heavy alcohol consumption, primary care physicians per 100,000, and mental health providers per 100,000 are obtained from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) and are provided at the county level. The BRFSS is a nationally representative annual telephone survey conducted by the CDC that assesses health-related risk behaviors throughout the United States. At more than 400,000 adult interviews every year, it is the largest health survey system in the world (CDC 2021).

Data on mental health providers per 100,000 was missing for 2010-2011. Missing value imputation was accomplished via Kalman smoothing. The heavy alcohol consumption variable is defined as consuming more than 4 (women) or 5 (men) alcoholic beverages on a single occasion in the past 30 days, or heavy drinking, defined as drinking more than 1 (women) or 2 (men) drinks per day on average. For the reporting poor/fair health variable, respondents were asked "In general, would you say that your health is excellent, very good, good, fair, or poor?". Respondents answering "poor" and "fair" were coded as poor/fair to create the poor/fair variable.

Multicollinearity

To check for multicollinearity, all variables were assessed for their Variance Inflation Factor (VIF). The Firearm composite had a VIF score of 4.9 while the no health insurance variable had a VIF score of 4.7. These were the highest VIF scores out of all the variables. A VIF score above 5 indicates moderate multicollinearity while a score above 10 indicates high multicollinearity. The results of the VIF assessment indicate that multicollinearity is not an issue in these analyses. A correlation matrix of all the variables can be observed below. It is imperative to observe the high correlation between the TM composite and the firearm composite of 0.73.

Figure 6. Correlation matrix of all covariates in the study



METHODS

The data are analyzed using fixed effects linear regression with robust standard errors clustering for DMA. Two-way fixed effects control for yearly and DMA fixed effects. Model specifications are consistent with previous area-level suicide studies that use two-way fixed effects with robust standard errors clustering by area (Brainerd 2001, Kunce & Anderson 2002, Neumayer 2003, Vitt et al. 2018, Anestis et al. 2017, Stack & Kposowa et al. 2016, Kposowa et al. 2016, Cylus et al. 2014). Model specification was supported via a Hausman test that indicated the appropriateness of using a fixed effects model over the random effects model.

The multivariate fixed effects model to be estimated is as follows:

$$SR_{ij} = \beta ST_{it} + \beta X_{ij} + \gamma_t + \alpha_i + \varepsilon_{ij}$$

The term SR is the male suicide rate per 100,000 in DMA *i* at time *t.* The variable of interest, Google search terms, is represented by vector ST. Vector X consists of the following control variables: marriage, divorce, male unemployment, veteran status, living alone, not having health insurance, reporting poor health, heavy alcohol consumption, primary care physicians per 100,000, and mental health providers per 100,000. Parameter γ represents time effects that account for unobserved national forces driving the male suicide rate. Parameter α represents unobserved DMA-specific fixed effects. The error term is represented by vector ε.

Fixed effect models control for potential omitted unobserved time-invariant area-specific factors that may affect the male suicide rate. Time-invariant factors that rarely undergo drastic changes through time are effectively controlled by the model. These factors include culture, climate, geography, elevation, religious composition, gender, and race. DMA-level factors that may contribute to suicide, such as a suicide culture, are controlled for as well.

Numerous studies cite the effect that economic cycles have on suicide rates (Wasserman 1984). Time-specific social currents such as recessions or major events that can affect suicide rates on a national level can provide misleading estimates. In efforts to not confound the estimates with an exogenous variable that may affect suicide trends, the specified model will control for such trends by controlling for yearly effects. It may be the case that a celebrity suicide may spur national increases in the male suicide rate, nevertheless such

instances are controlled for by the yearly fixed effects. Standard errors are clustered by DMA to address serial correlation which biases standard errors. Analysis is conducted using the fixest package in R.

RESULTS

Several models were estimated. First, bivariate linear regressions were run with each variable as the independent variable and the male suicide rate as the dependent variable. None of these models included fixed effects. Robust standard errors were clustered at the DMA level.

Bivariate models are run to demonstrate each variable’s independent relationship with the male suicide rate, without controlling for other unobserved confounders. This allows us to verify each variable’s association with the male suicide rate, however this leaves room for unobserved confounders.

The results of these bivariate models are shown in table 1 below. The TM search term composite is significant at the 0.001 level in the expected direction, and produces an r-squared of 0.09. The firearm search term composite is significant at the 0.001 level in the expected direction, and produces a big r-squared of 0.18. Divorce and being of veteran status produce the largest r-squared of 0.33 and 0.20 respectively. They are both in the expected direction. The largest coefficient is produced by divorce (2.646).

Table 1. Bivariate linear regression. Dependent variable: male suicide rate (2010-2016)

Linear Regression (Bivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE	R2
TM Search Term Composite	**0.2974***	0.0411	0.09
Firearm Search Term Composite	**0.2003***	0.0178	0.18
Married	**0.5938***	0.1015	0.10
Divorce	**2.646***	0.2221	0.33
Male Unemployment	-0.2164	0.1765	0.00
Veteran	**1.388***	0.1551	0.20
Live Alone	**1.463***	0.2560	0.11
No Health Insurance	0.0344	0.0896	0.00
Alcohol	-0.0905	0.0947	0.00
Poor Health	-0.1951	0.1252	0.01
Physicians per Capita	-0.0200	0.0119	0.00
Mental Health Providers per Capita	**0.0085***	0.0034	0.01
Fixed-Effects:	None		
SE: Clustered	Robust / Clustered by DMA		
Observations:	1,463		

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Next, three multivariate linear regressions were run with the male suicide rate as the dependent variable. These models do not control for any fixed effects. Robust standard errors were clustered at the DMA level. The results of these multivariate models are shown in tables 2, 3, and 4 below. These models are run to observe the relationship these covariates have with the male suicide rate, and to observe which associations are no longer significant as a result of controlling for covariates. It is important to note that these models do not control for other unobserved confounders. Unobserved confounders will be accounted for in the fixed effects models.

Three models were run. The first one (table 2) includes both the TM search term composite and the firearm search term composite. In this model, only the firearm search term composite is significant at the 0.001 level in the expected direction. In the second model (table 3), the firearm search term composite is omitted from the model. In this model, the TM search term composite is significant at the 0.001 level in the expected direction. In the third model (table 4), the TM search term is omitted from the model. In this model, the firearm search term composite is significant at the 0.001 level in the expected direction. The model with the best fit is the one with only the firearm search term composite, which produces an r-squared of 0.514, compared to one with only the TM search term composite which produces an r-squared of 0.50.

It is notable that divorce and being of veteran status remain significant at the 0.001 level in the expected direction. Divorce produces some of the highest coefficients thus far. Another note is that not having insurance becomes significant at the 0.001 level in the expected direction in the multivariate analyses. It is important to note that both the TM and the firearm search term composites remain significant when placed alone in the multivariate models. This demonstrates that these search term composites are significantly associated with the male suicide rate independent of the major suicide predictors from the literature. Therefore, these Google search term composites can add to what is already known about male suicide. Lastly, it must be noted that when both composites are included in the same model (table 2), only the firearm composite is significant and the TM composite ceases to be significant.

Table 2. Multivariate linear regression. Dependent variable: male suicide rate (2010-2016). Includes both Traditional Masculinity and Firearm search composites.

Linear Regression (Multivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE
TM Search Term Composite	0.0174	0.0435
Firearm Search Term Composite	**0.0971*** **	0.0220
Married	**0.4505*** **	0.0793
Divorce	**1.5747*** **	0.2175
Male Unemployment	0.0202	0.1043
Veteran	**0.4885*** **	0.1217
Live Alone	**0.6788* **	0.2984
No Health Insurance	**0.1891*** **	0.0699
Alcohol	0.0092	0.1017
Poor Health	-0.2053`	0.1192
Physicians per Capita	0.0052	0.0114
Mental Health Providers per Capita	**0.0086*** **	0.0021
Fixed-Effects:	None	
SE:	Robust / Clustered by DMA	
Observations:	1,463	
R2	0.514	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Table 3. Multivariate linear regression. Dependent variable: male suicide rate (2010-2016). Includes only Traditional Masculinity search composite.

Linear Regression (Multivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE
TM Search Term Composite	**0.1209*** **	0.0352
Firearm Search Term Composite	-	-
Married	**0.4702*** **	0.0774
Divorce	**1.805*** **	0.2018
Male Unemployment	-0.0593	0.1064
Veteran	**0.4526*** **	0.1174
Live Alone	**0.7353* **	0.3069
No Health Insurance	**0.1972*** **	0.0705
Alcohol	0.0107	0.1043
Poor Health	-0.1938	0.1223
Physicians per Capita	-0.0091	0.0102
Mental Health Providers per Capita	**0.0106*** **	0.0020
Fixed-Effects:	None	
SE:	Robust / Clustered by DMA	
Observations:	1,463	
R2	0.50	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Table 4. Multivariate linear regression. Dependent variable: male suicide rate (2010-2016). Includes only Firearm search composite.

Linear Regression (Multivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE
TM Search Term Composite	-	-
Firearm Search Term Composite	**0.1028*** **	0.0184
Married	**0.4485*** **	0.0789
Divorce	**1.5690*** **	0.2156
Male Unemployment	0.0214	0.1050
Veteran	**0.4944*** **	0.1181
Live Alone	**0.6869* **	0.2998
No Health Insurance	**0.1880*** **	0.0694
Alcohol	0.0093	0.1016
Poor Health	-0.2019	0.1177
Physicians per Capita	0.0046	0.0108
Mental Health Providers per Capita	**0.0086*** **	0.0021
Fixed-Effects:	None	
SE:	Robust / Clustered by DMA	
Observations:	1,463	
R2	0.514	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Bivariate linear regressions were run with each variable as the independent variable and the male suicide rate as the dependent variable. These models include both yearly and DMA-level fixed effects. Robust standard errors were clustered at the DMA level. These bivariate models are run to demonstrate each variable's independent relationship with the male suicide rate, while controlling for unobserved confounders. This allows us to verify each variable's association with the male suicide rate.

The results of these bivariate models are shown in table 5 below. In these models, the fixed effects control for all time-invariant factors. Essentially, all variables are controlled for due to a lack of variance over time. This implies that previously significant variables such as divorce and being of veteran status are generally steady factors within

DMAs over the period of 2010 through 2016. The same appears to be the case for both search term composites. These findings suggest that TM-related searches and firearm-related searches tend to be concentrated in the same areas and their interest is steady through time. It is worth noting that living alone is significant at the 0.10 level and produces the largest r-squared out of any variable in these analyses.

Table 5. Two-way fixed effects bivariate linear regression. Dependent variable: male suicide rate (2010-2016).

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Bivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE	R2
TM Search Term Composite	0.0124	0.0428	0
Firearm Search Term Composite	0.0531	0.0498	0.005
Married	-0.2878	0.3881	0
Divorce	1.0982	0.7715	0.005
Male Unemployment	0.0838	0.2442	0
Veteran	1.0598	0.8163	0.003
Live Alone	1.6193`	0.9711	0.008
No Health Insurance	-0.0935	0.0979	0.001
Alcohol	-0.0663	0.1378	0
Poor Health	-0.1921	0.1203	0.001
Physicians per Capita	0.0121	0.0100	0
Mental Health Providers per Capita	0.0011	0.0024	0
Fixed-Effects:	Year & DMA		
SE: Clustered	Robust / Clustered by DMA		
Observations:	1,463		

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Lastly, three multivariate linear regression models were run with the male suicide rate as the dependent variable. The first model includes both search term composites (table 6). The second model only includes the TM search term composite (table 7). The third model only includes the firearm search term composite (table 8). These models include both yearly and DMA fixed effects. Robust standard errors are clustered at the DMA level. The results of these multivariate models are shown in tables 6, 7, and 8 below.

These models are run to observe the relationship the covariates have with the male suicide rate, and to see which associations are no longer significant as a result of controlling for the covariates. Additionally, these models allow us to observe if any variables become significant after controlling for unobserved confounders.

None of these models produce any significant findings. Again, while some of these variables were prominently significant before implementing the two-way fixed effects, like divorce and veteran, their bearing on the male suicide rate has been effectively controlled for due to their lack of variance over the seven year period.

Table 6. Two-way fixed effects multivariate linear regression. Dependent variable: male suicide rate (2010-2016). Includes both Traditional Masculinity and Firearm search composites.

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Multivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE
TM Search Term Composite	0.0014	0.0389
Firearm Seach Term Composite	0.0459	0.0461
Married	0.2543	0.4289
Divorce	0.9277	0.8112
Male Unemployment	0.1173	0.2075
Veteran	1.1360	0.7651
Live Alone	1.5577	0.9766
No Health Insurance	-0.1277	0.1003
Alcohol	-0.0816	0.1263
Poor Health	-0.2323	0.1491
Physicians per Capita	0.0099	0.0101
Mental Health Providers per Capita	0.0018	0.0025
Fixed-Effects:	Year & DMA	
SE:	Robust / Clustered by DMA	
Observations:	1,463	
R2	0.025	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Table 7. Two-way fixed effects multivariate linear regression. Dependent variable: male suicide rate (2010-2016). Includes only Traditional Masculinity search composite.

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Multivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE
TM Search Term Composite	0.0150	0.0429
Firearm Seach Term Composite	-	-
Married	0.2473	0.4299
Divorce	1.0154	0.8707
Male Unemployment	0.1345	0.2188
Veteran	1.1899	0.8056
Live Alone	1.5573	1.0120
No Health Insurance	-0.1448	0.0993
Alcohol	-0.0554	0.1330
Poor Health	-0.2241	0.1434
Physicians per Capita	0.0114	0.0104
Mental Health Providers per Capita	0.0019	0.0025
Fixed-Effects:	Year & DMA	
SE:	Robust / Clustered by DMA	
Observations:	1,463	
R2	0.022	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

Table 8. Two-way fixed effects multivariate linear regression. Dependent variable: male suicide rate (2010-2016). Includes only Firearm search composite.

Fixed Effects Linear Regression with DMA and Yearly Fixed Effects (Multivariate)

Dependent Variable: All-Cause Male Suicide Rate (2010-2016)

	Coefficient	SE
TM Search Term Composite	-	-
Firearm Seach Term Composite	0.0461	0.0466
Married	0.2537	0.4297
Divorce	0.9276	0.8105
Male Unemployment	0.1170	0.2062
Veteran	1.1357	0.7643
Live Alone	1.5566	0.9804
No Health Insurance	-0.1279	0.0990
Alcohol	-0.0815	0.1259
Poor Health	-0.2325	0.1473
Physicians per Capita	0.0099	0.0100
Mental Health Providers per Capita	0.0018	0.0025
Fixed-Effects:	Year & DMA	
SE:	Robust / Clustered by DMA	
Observations:	1,463	
R2	0.025	

***p<0.001, **p<0.01, *p<0.05, `p<0.1

These results demonstrate that both search term composites are significantly associated with the male suicide rate, even after controlling for the major male suicide predictors from the literature. This demonstrates that these search term composites are capable of revealing more than what is already known from previous studies. The fact that the search term composites are no longer significant in the fixed effects models suggests that these composites may be capturing an area's “culture”. That is, the areas with the highest levels of firearm-interest happen to be the same areas, and the level of interest remains stable throughout the entire time frame of 2010 to 2016. The same can be said about TM-culture; it appears to be stable in the same areas over the period of the analyses.

Lagged Model

While the models discussed above assess the variables and the suicide rate on a yearly basis, this approach can bring up questions of causality. Therefore, all independent variables were lagged by one year and regressed on next year's male suicide rate. For example, the TM-composite of 2015 was regressed on the 2016 male suicide rate. To conduct this analysis, the 2009 entries for all independent variables were obtained. Every bivariate model was run with the male suicide rate as the dependent variable, and the one-year lag of each covariate as the independent variable, while controlling for the one-year lag of the dependent variable, the male suicide rate. No variables were significant in these bivariate analyses. Next, a multivariate model was run with the male suicide rate as the dependent variable and the one-year lags of all covariates, including the one-year lag of the dependent variable, the male suicide rate. No variables were significant in the multivariate analysis. The fact that there are no significant findings indicates to the reduced ability in using the one-year lag of TM-related and firearm-related search terms in attempting to predict next year's male suicide rates. While the year-by-year analyses provide promising results, this study cannot assert that the TM composite, nor the firearm composite, has any significant predictive power.

Robustness Check

It is possible that the Google search composites may be confounding with an unobserved aspect of an area rather than a specific phenomenon related to male suicide. To test this possibility, we examined other possible health outcomes, cancer mortality and heart attack mortality, as dependent variables. If the Google search composites are significantly associated with such outcomes, then this study cannot argue on the reliability of Google data and its association with the male suicide rate.

Cancer mortality and heart attack mortality rate data was obtained from the CDC compressed mortality file. Cancer mortality is defined by ICD-10 codes C00-C97. Heart attack mortality is defined by ICD-10 codes I210-I219. Cancer and heart attack mortality rates were aggregated by DMA and merged to the final dataset.

The TM composite was tested in a bivariate fixed effects model controlling for DMA and yearly fixed effects. The TM composite did not have an association with cancer mortality ($t = -1.54$). The TM composite did not have an association with heart attack mortality ($t = 0.75$).

The firearm composite was tested in a bivariate fixed effects model controlling for DMA and yearly fixed effects. The firearm composite did not have an association with cancer mortality ($t = 0.44$). The firearm composite did not have an association with heart attack mortality ($t = 0.76$).

Therefore, these analyses strengthen the assertion that Google search term data may reveal profound social currents that dispose individuals towards succumbing to the most personal of all choices, and that such associations are not spurious in nature.

DISCUSSION

The discussion section summarizes the findings from the result sections, focusing on the TM and firearm search term composite, and arguing that Google search data can add to what is already known from the literature. It touches on the strengths of implementing Google search composites opposed to using single search terms. The discussion finishes by citing the importance of using Google search data and other internet sources that can provide researchers with rich information on human attitudes, behaviors, and inclinations.

The considerable consistency and regularity of suicide rates over time demonstrate that each geographical location has a definite aptitude for suicide. This aptitude is a function of the shared beliefs, attitudes, sentiments, and culture that hold people together. An individual succumbs to an anomic or stressful event such as an occupational loss, relationship breakdown, or bankruptcy, not because of the event itself, but rather how conducive the culture is towards suicide (Durkheim 1951[1897]). This study argues that cultures are measurable via Google search data, and if culture can be measured, then its subsequent behaviors, thoughts, and inclinations can also be measured. In this instance, the culture of Traditional Masculinity and firearms can be said to be associated with rate of male suicide.

The analyses demonstrate that both search term composites are significantly associated with the male suicide rate in the linear regression models that do not include fixed effects (tables 1- 4). It must also be noted that both composites remain significant far controlling for the major male suicide predictors from the literature. This finding indicates that Google search term data can add to what is already known in the study of male suicide. It can further be added that Google searches provide us with another way of observing a society, an area's cultural interests, and its subsequent behavior.

It must be noted however that when the composites are analyzed when controlling for unobserved confounding, via the fixed effects models (tables 5-8), the composites are no longer significant which indicates that there may be an time-varying unobserved factor driving the male suicide rates. Other possible time-varying unobserved factors that may drive the male suicide rate include long-term unemployment (Kposowa et al. 2019), firearm ownership, or a social phenomena such as Case & Deaton's (2020) concept of despair, that requires further study. It is also possible that both search term composites are accurately measuring male suicides, however since both Traditional Masculinity and firearms are cultures, and are likely to be stable geographically and through time, then both composites are effectively controlled for in the fixed effects models.

It is also imperative to note that both composites are correlated with each other. Masculinity and the traditional masculinity ideology has an association with firearm ownership (Borgogna et al. 2022). There is overlap between subscribing to TM culture and owning a firearm and therefore, both composites may be part of a bigger social construct that is not captured by these analyses. It is notable that the firearm composite performed well when compared to the TM composite. In the bivariate models (table 1), the firearm composite produced an r-squared of 0.18, which is considerably high in social analyses, while the TM composite produced an r-squared of 0.09, still relatively high. When controlling for other covariates (table 3-4), the TM composite and the firearm composite produce coefficients of 0.12 and 0.10 respectively. When interpreted, a ten-point increase in the search term frequency for Traditional Masculinity-related topics is associated with 1.2 more male suicide deaths per 100,000 population. For an area of one million people, this equates to about 6 extra deaths, or roughly 150 years of potential life lost. A ten-point increase in the search term frequency for firearm-related topics is associated with 1.0 more male suicide deaths per 100,000 population. For an area of one million people, this equates to about 5 extra deaths, or roughly 125 years of potential life lost. Years of potential life lost is calculated by subtracting the age of suicide by the standard life expectancy of 75, then taking the conservative assumption that all death occurred around the age of 50 (CDC 2018).

Using principal component analysis aiming to create a culture composite was successful in this study. The principal component analysis analyzed over 48 TM-related search terms and over 44 firearm-related search terms. The number of search terms analyzed by this study exceeds the number from previous research. This success of the principal component analysis gives some credence to the idea that search term composites may assist in assessing social phenomena better than single terms. The logic behind this is that an area might report a spike in searches for “guns” because of a school shooting on the news, however an area that reports a spike in searches for “guns” and brands of guns and gun accessories is far more likely to possess firearm culture. Therefore, Google search composites may provide a more true and holistic lens into an area’s culture and inclinations.

No association was observed for the one-year lag of the two search term composites, indicating that they do not possess predictive power over a year of lag. A reason for this null finding may be the fact that the composites capture culture, a salient social factor that is geographically and time-invariant. Therefore if the levels of culture remain the same over time, lagged analyses would not be a good logical idea. A descriptive analysis of the DMAs with the highest and lowest composite scores show that DMAs high and low in TM or firearm culture remain stable over the course of this study (2010-2016). Therefore, lagged models may perform better when behaviors such as home-buying interest are assessed. However, in this case, since most search terms in this study are descriptive, they better encapsulate a culture, where lagged models make less sense.

The finding of divorce and veteran status being significantly associated with male suicide supports previous findings. Divorce can be defined as an anomic event that can trigger negative emotions, create scenarios of normlessness, and reduce one's social integration. These pathways are conducive to suicide, and perhaps divorce may be more difficult to endure for males who subscribe to TM culture. Additionally, veterans are about 50 percent higher risk of suicide than those who have not served in the military. Since 2006, there has been an 80 percent increase in suicide among veterans (Stop Soldier Suicide 2022).

To summarize, the results of the models discussed above demonstrate that the variable of interest, the TM and firearm search term composites, remain robust throughout the linear regression models without fixed effects. They remain significant after controlling for the major predictors of male suicide from the literature. As previous studies using Google search terms have been able to assess numerous health outcomes, including herpes, plague, cancer, whooping cough, influenza, asthma, Lyme disease, syphilis, HIV, depression, obesity, drug use, and many more, (Mavragani & Ochoa, 2019, Arora et al. 2019, Jun et al. 2018, Brodeur et al. 2021), this study has successfully used two cultural search term composites to assess the male suicide rate at the DMA-level over the period of 2010 to 2016.

LIMITATIONS

Not all Google searches for TM-related or firearm-related content are necessarily made by people subscribing to such cultures. Nonetheless, an area conducting a more than average search volume for this type of content can be said to be reflecting the overall social environment of an area. For example, Googling for the Smith & Wesson M&P model firearm does not necessarily mean one owns a firearm or even has the intent to purchase one, however they might have heard about the firearm in a conversation or on an advertisement. Therefore, Google searches for TM and firearm-related content reflect situations found in one's surrounding social environment, and that is exactly what this study attempts to capture.

Religiosity is a sociological construct that should have a negative effect on suicide, and therefore protect against suicide (Durkheim 1951[1897]). However, data on area-level religiosity can only be found at the state-level through the General Social Survey and therefore data at the DMA-level is not available. Using Google Trends data to measure area-level religiosity has not been thoroughly tested and is a potentially new area for exploration. Thus far, there has been one study investigating religiosity via Google search data with promising results (Yeung 2019).

Also, given that roughly half of all suicides are completed via firearm, a major factor to include in the analyses is firearm ownership. However, data on this is elusive and only available at the state-level. The General Social Survey stopped asking about firearm ownership two decades ago. Currently, the most reliable data on firearm ownership is from the FBI's National Instant Criminal Background Check System (NICS) and is only available at the state level. Several studies have shown the NICS's viability in assessing the suicide rate at the state-level (Lang 2013, Vitt et al. 2018). However, obtaining reliable firearm ownership data below the state-level remains a challenge. This study shows that since firearm ownership data is unavailable, a viable proxy could be the extent to which an area demonstrates interest in firearms and firearm-related content on Google. This can be tested and verified thoroughly using the state-level NICS data.

Lastly, most of the control variables are not gender specific except for male unemployment. Obtaining the male data on veterans, living alone, not having health insurance, heavy alcohol consumption, and poor health would strengthen the analyses however only combined male and female data was available from the ACS.

POLICY IMPLICATIONS

These findings produce several policy implications. Suicide rates in the US have been rising dramatically in the last few years. This analysis points that some of the major factors behind male suicide are divorce, being of veteran status, and subscribing to TM and/or firearm culture. It is imperative that mental health clinicians are vigilant of traits or characteristics that suggest an individual subscribes to such cultures, largely because it increases their risk of completing suicide.

Regarding cultural adherence to TM, areas that frequently Google for these topics should be watched closely by the CDC. In fact, because culture is relatively stable, these areas can already be observed using the composites created by this study. The DMAs demonstrating the highest levels of TM and firearm culture are also the areas that pose the greatest risk of suicide for males. Knowledge regarding high TM and firearm culture areas can inform policy and health initiatives, such as where to allocate suicide prevention resources. Furthermore, mental health resources should be allocated to areas that have a greater proportion of inhabitants that are of veteran status.

ACADEMIC CONTRIBUTIONS

This project advances epidemiology and suicidology research by implementing an innovative methodology that is seldom used. Additionally, this project advances the emerging field of digital epidemiology, whose studies show that our online digital traces leave behind clues about our health. Previous studies have validated this methodology at the national and state-level for suicide, however no previous studies have done these analyses at the DMA-level. This study contributes to the study of suicide by being the first to do these analyses at the DMA-level, which is the smallest geographic tract available in Google Trends.

This project advances sociological and anthropological inquiry by attempting to capture geographic variation in culture, and then tying such culture to a hard outcome, male suicide. Previous research has used interviews and surveys to assess the extent that an individual or an area subscribes to a culture. However, this study demonstrates that Google search data, a low cost alternative, can also add to what is already known about areas and the cultures they subscribe to. Additionally it allows for the opportunity to assess if such cultures can be tied to hard health outcomes, such as how this project shows a type of masculine culture and firearm culture, measured by Google searches, can be associated with an irreversible health outcome.

Lastly, this project revives the long sociological tradition of studying suicide, albeit through a more modern and innovative lens. This study introduces an emerging and verified methodology to sociology. Sociologists have not identified innovative and sociologically interesting approaches to suicide that are available in other disciplines (Wray et al. 2011). Current sociological research lacks up-to-date methods and instead focuses heavily on surveys and interviews. Since 1980, of the 30,000 academic articles on suicide, only 1.3 percent were categorized as sociological (Wray et al. 2011). Given the dearth of recent sociological research on suicide, this project fills many gaps in the methods of studying suicide, and should inspire sociologist to use the digital remnants left being by our society.

CONCLUSION

While previous studies have attempted to study the association between suicidal Google searches and the suicide rate, this study captures the geographic variation in culture and then ties it to the male suicide rate, using a solid theory. This study finds that Google searches are able to capture area level culture, in this case Traditional Masculinity culture and firearm culture, and find a significant association with such cultures and the male suicide rate. This study opens the door for sociologists and anthropologists to assess culture via our online traces.

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews. Computational Statistics*, *2*(4), 433–459.

Adam-Troian, J., & Arciszewski, T. (2020). Absolutist Words From Search Volume Data Predict State-Level Suicide Rates in the United States. *Clinical Psychological Science*, *8*(4), 788–793.

Alden, H. L., & Parker, K. F. (2005). Gender Role Ideology, Homophobia and Hate Crime: Linking Attitudes to Macro-Level Anti-Gay and Lesbian Hate Crimes. *Deviant Behavior*, *26*(4), 321–343.

Andrés, A. R. (2005). Income inequality, unemployment, and suicide: a panel data analysis of 15 European countries. *Applied Economics*, *37*(4), 439–451.

Anestis, M. D., & Houtsma, C. (2018). The Association Between Gun Ownership and Statewide Overall Suicide Rates. *Suicide & Life-Threatening Behavior*, *48*(2), 204–217.

Anestis, M. D., Selby, E. A., & Butterworth, S. E. (2017). Rising longitudinal trajectories in suicide rates: The role of firearm suicide rates and firearm legislation. *Preventive Medicine*, *100*, 159–166.

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, *58*(2), 277–297.

Arendt, F. (2021). The Opioid-Overdose Crisis and Fentanyl: The Role of Online Information Seeking via Internet Search Engines. *Health Communication*, *36*(10), 1148–1154.

Arora, V. S., McKee, M., & Stuckler, D. (2019). Google Trends: Opportunities and limitations in health and health policy research. *Health Policy* , *123*(3), 338–341.

Ayers, J. W., Althouse, B. M., Allem, J.-P., Rosenquist, J. N., & Ford, D. E. (2013). Seasonality in seeking mental health information on Google. *American Journal of Preventive Medicine*, *44*(5), 520–525.

Beck, N. L., Katz, J. N., & Mignozzetti, U. G. (2014). Of Nickell Bias and its Cures: Comment on Gaibulloev, Sandler, and Sul. *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association*, *22*(2), 274–278.

Bertolote, J. M., & Fleischmann, A. (2002). Suicide and psychiatric diagnosis: a worldwide perspective. *World Psychiatry: Official Journal of the World Psychiatric Association* , *1*(3), 181–185.

Biddle, L., Derges, J., Mars, B., Heron, J., Donovan, J. L., Potokar, J., Piper, M., Wyllie, C., & Gunnell, D. (2016). Suicide and the Internet: Changes in the accessibility of suicide-related information between 2007 and 2014. *Journal of Affective Disorders*, *190*, 370–375.

Borgogna, N. C., McDermott, R. C., & Brasil, K. M. (2022). The precarious masculinity of firearm ownership. *Psychology of Men & Masculinities*, *23*(2), 173–182.

Borowsky, S. J., Rubenstein, L. V., Meredith, L. S., Camp, P., Jackson-Triche, M., & Wells, K. B. (2000). Who is at risk of nondetection of mental health problems in primary care? *Journal of General Internal Medicine*, *15*(6), 381–388.

Brainerd, E. (2001). Economic reform and mortality in the former Soviet Union: A study of the suicide epidemic in the 1990s. *European Economic Review*, *45*(4), 1007–1019.

Briggs, J. T., & Tabarrok, A. (2014). Firearms and suicides in US states. *International Review of Law and Economics*, *37*, 180–188.

Brodeur, A., Clark, A. E., Fleche, S., & Powdthavee, N. (2021). COVID-19, lockdowns and well-being: Evidence from Google Trends. *Journal of Public Economics*, *193*, 104346.

Burnes, B., & Bargal, D. (2017). Kurt Lewin: 70 Years on. *Journal of Change Management*, *17*(2), 91–100.

Burnes, B., & Cooke, B. (2012). Kurt lewin's field theory: A review and re-evaluation. *International Journal of Management Reviews*. https://doi.org/10.1111/j.1468-2370.2012.00348.x

Campo, D. S., Gussler, J. W., Sue, A., Skums, P., & Khudyakov, Y. (2020). Accurate spatiotemporal mapping of drug overdose deaths by machine learning of drug-related web-searches. *PloS One*, *15*(12), e0243622.

Canetto, S. S. (1997). Meanings of gender and suicidal behavior during adolescence. *Suicide & Life-Threatening Behavior*, *27*(4), 339–351.

Capron, D. W., Andel, R., Voracek, M., Till, B., Niederkrotenthaler, T., Bauer, B. W., Anestis, M. D., & Tran, U. S. (2021). Time-series analyses of firearm-related Google searches and U.S. suicide rates 2004-2016. *Suicide & Life-Threatening Behavior*, *51*(3), 554–563.

Case, A., & Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(49), 15078–15083.

Case, A., & Deaton, A. (2021). Life expectancy in adulthood is falling for those without a BA degree, but as educational gaps have widened, racial gaps have narrowed. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(11). https://doi.org/10.1073/pnas.2024777118

Case, A., Deaton, A., & Stone, A. A. (2020). Decoding the mystery of American pain reveals a warning for the future. *The National Academy of Sciences*, *117*(40), 24785–24789.

Cerel, J., van de Venne, J. G., Moore, M. M., Maple, M. J., Flaherty, C., & Brown, M. M. (2015). Veteran exposure to suicide: Prevalence and correlates. *Journal of Affective Disorders*, *179*, 82–87.

Chae, D. H., Clouston, S., Hatzenbuehler, M. L., Kramer, M. R., Cooper, H. L. F., Wilson, S. M., Stephens-Davidowitz, S. I., Gold, R. S., & Link, B. G. (2015). Association between an Internet-Based Measure of Area Racism and Black Mortality. *PloS One*, *10*(4), e0122963.

Chae, D. H., Clouston, S., Martz, C. D., Hatzenbuehler, M. L., Cooper, H. L. F., Turpin, R., Stephens-Davidowitz, S., & Kramer, M. R. (2018). Area racism and birth outcomes among Blacks in the United States. *Social Science & Medicine*, *199*, 49–55.

Chamie, J. (2021, July 19). *Living alone in America*. https://thehill.com/opinion/finance/563786-living-alone-in-america/

Chan, J. (2019). The effect of college education on intolerance: evidence from Google search data. *Applied Economics Letters*, *26*(2), 83–86.

Chang, S.-S., Kwok, S. S. M., Cheng, Q., Yip, P. S. F., & Chen, Y.-Y. (2015). The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis. *Social Psychiatry and Psychiatric Epidemiology*, *50*(9), 1451–1461.

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *The Economic Record*, *88*, 2–9.

Chykina, V., & Crabtree, C. (2018). Using Google Trends to Measure Issue Salience for Hard-to-Survey Populations. *Socius*, *4*, 2378023118760414.

Cohany, S. R., & Sok, E. (2007). Trends in labor force participation of married mothers of infants. *Monthly Labor Review / U.S. Department of Labor, Bureau of Labor Statistics*, *130*, 9.

Coleman, D. (2015). Traditional Masculinity as a Risk Factor for Suicidal Ideation: Cross-Sectional and Prospective Evidence from a Study of Young Adults. *Archives of Suicide Research: Official Journal of the International Academy for Suicide Research*, *19*(3), 366–384.

Coleman, D., & Casey, J. T. (2011). The Social Nature of Male Suicide: A New Analytic Model. *International Journal of Men's Health*. https://doi.org/10.3149/jmh.1003.240

Coleman, D., Feigelman, W., & Rosen, Z. (2020). Association of High Traditional Masculinity and Risk of Suicide Death: Secondary Analysis of the Add Health Study. *JAMA Psychiatry* , *77*(4), 435–437.

Cousins, H. C., Cousins, C. C., Harris, A., & Pasquale, L. R. (2020). Regional Infoveillance of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns. *Journal of Medical Internet Research*, *22*(7), e19483.

Curtis, D. S., Washburn, T., Lee, H., Smith, K. R., Kim, J., Martz, C. D., Kramer, M. R., & Chae, D. H. (2021). Highly public anti-Black violence is associated with poor mental health days for Black Americans. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(17). https://doi.org/10.1073/pnas.2019624118

Cylus, J., Glymour, M. M., & Avendano, M. (2014). Do generous unemployment benefit programs reduce suicide rates? A state fixed-effect analysis covering 1968-2008. *American Journal of Epidemiology*, *180*(1), 45–52.

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, *33*(4), 801–816.

Davey, Z., Schifano, F., Corazza, O., Deluca, P., & Psychonaut Web Mapping Group. (2012). e-Psychonauts: conducting research in online drug forum communities. *Journal of Mental Health* , *21*(4), 386–394.

Davis, C. S., Lieberman, A. J., Hernandez-Delgado, H., & Suba, C. (2019). Laws limiting the prescribing or dispensing of opioids for acute pain in the United States: A national systematic legal review. *Drug and Alcohol Dependence*, *194*, 166–172.

Dean, A., & Kimmel, S. (2019). Free trade and opioid overdose death in the United States. *SSM - Population Health*, *8*, 100409.

Definitions for Years of Potential Life Lost - Fatal Injury Help Menu. (2018, July 24). https://www.cdc.gov/injury/wisqars/fatal_help/definitions_ypll.html

Deisenhammer, E. A., Ing, C.-M., Strauss, R., Kemmler, G., Hinterhuber, H., & Weiss, E. M. (2009). The duration of the suicidal process: how much time is left for intervention between consideration and accomplishment of a suicide attempt? *The Journal of Clinical Psychiatry*, *70*(1), 19–24.

DeWeerdt, S. (2019). Tracing the US opioid crisis to its roots. *Nature*, *573*(7773), S10–S12.

Dowell, D., Zhang, K., Noonan, R. K., & Hockenberry, J. M. (2016). Mandatory Provider Review And Pain Clinic Laws Reduce The Amounts Of Opioids Prescribed And Overdose Death Rates. *Health Affairs* , *35*(10), 1876–1883.

Drew, D. A., Nguyen, L. H., Steves, C. J., Menni, C., Freydin, M., Varsavsky, T., Sudre, C. H., Cardoso, M. J., Ourselin, S., Wolf, J., Spector, T. D., Chan, A. T., & COPE Consortium. (2020). Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science*, *368*(6497), 1362–1367.

Durkheim, E. (1951). Suicide: A study in sociology (JA Spaulding & G. Simpson, trans.). *Glencoe, IL: Free Press. (Original Work Published 1897)*.

Eriksen, C. U., Konradsen, F., & Vildekilde, T. (2021). Searching Online for Methods of Suicide. *Crisis*, *42*(2), 136–143.

Fox, S., & Rainie, H. (2014). *The Web at 25 in the US*. Pew Research Center [Internet & American Life Project].

Ghertner, R., & Groves, L. (2018). Economic Opportunity and the Opioid Crisis: Geographic and Economic Trends. *US Department of Health and Human Services*, 1–22.

Gimbrone, C., Rutherford, C., Kandula, S., Martínez-Alés, G., Shaman, J., Olfson, M., Gould, M. S., Pei, S., Galanti, M., & Keyes, K. M. (2021). Associations between COVID-19 mobility restrictions and economic, mental health, and suicide-related concerns in the US using cellular phone GPS and Google search volume data. *PloS One*, *16*(12), e0260931.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014.

Girard, C. (1993). Age, Gender, and Suicide: A Cross-National Analysis. *American Sociological Review*, *58*(4), 553–574.

Gregory, M. (2012). Masculinity and homicide-suicide. *International Journal of Law, Crime and Justice*, *40*(3), 133–151.

Guerra, C. (2019). *Internet Search Data: A Better Proxy for Regional Suicidality* [Masters in Sociology]. UC Riverside.

Gunn, J. F., 3rd, & Lester, D. (2013). Using google searches on the internet to monitor suicidal behavior. *Journal of Affective Disorders*, *148*(2-3), 411–412.

Gunnell, D., Derges, J., Chang, S.-S., & Biddle, L. (2015). Searching for Suicide Methods: Accessibility of Information About Helium as a Method of Suicide on the Internet. *Crisis*, *36*(5), 325–331.

Hagihara, A., Miyazaki, S., & Abe, T. (2012). Internet suicide searches and the incidence of suicide in young people in Japan. *European Archives of Psychiatry and Clinical Neuroscience*, *262*(1), 39–46.

Hall, J., Lo, F., Saha, S., Vaidyanathan, A., & Hess, J. (2020). Internet searches offer insight into early-season pollen patterns in observation-free zones. *Scientific Reports*, *10*(1), 11334.

Harper, S., Riddell, C. A., & King, N. B. (2021). Declining Life Expectancy in the United States: Missing the Trees for the Forest. *Annual Review of Public Health*, *42*, 381–403.

Hassett, A. L., Aquino, J. K., & Ilgen, M. A. (2014). The risk of suicide mortality in chronic pain patients. *Current Pain and Headache Reports*, *18*(8), 436.

Hatzenbuehler, M. L., Keyes, K. M., & Hasin, D. S. (2009). State-level policies and psychiatric morbidity in lesbian, gay, and bisexual populations. *American Journal of Public Health*, *99*(12), 2275–2281.

Hawton, K., Casañas I Comabella, C., Haw, C., & Saunders, K. (2013). Risk factors for suicide in individuals with depression: a systematic review. *Journal of Affective Disorders*, *147*(1-3), 17–28.

Hayes, J. A., & Mahalik, J. R. (2000). Gender role conflict and psychological distress in male counseling center clients. *Psychology of Men & Masculinity*, *1*(2), 116–125.

Hooley, J. M., Franklin, J. C., & Nock, M. K. (2014). Chronic pain and suicide: understanding the association. *Current Pain and Headache Reports*, *18*(8), 435.

Houle, J., Mishara, B. L., & Chagnon, F. (2008). An empirical test of a mediation model of the impact of the traditional male gender role on suicidal behavior in men. *Journal of Affective Disorders*, *107*(1-3), 37–43.

How Many Guns are in the US? (Gun Ownership Statistics). (2021, October 22). American Gun Facts. https://americangunfacts.com/gun-ownership-statistics/

Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing*, *285*, 188–195.

Jeong, J., Shin, S. D., Kim, H., Hong, Y. C., Hwang, S. S., & Lee, E. J. (2012). The effects of celebrity suicide on copycat suicide attempt: a multi-center observational study. *Social Psychiatry and Psychiatric Epidemiology*, *47*(6), 957–965.

Joiner, T. E., & The Robert O Lawton Distinguished Professor Thomas Joiner. (2005). *Why People Die by Suicide*. Harvard University Press.

Joiner, T. E., Van Orden, K. A., Witte, T. K., Selby, E. A., Ribeiro, J. D., Lewis, R., & Rudd, M. D. (2009). Main predictions of the interpersonal-psychological theory of suicidal behavior: empirical tests in two samples of young adults. *Journal of Abnormal Psychology*, *118*(3), 634–646.

Ju Ji, N., Young Lee, W., Seok Noh, M., & Yip, P. S. F. (2014). The impact of indiscriminate media coverage of a celebrity suicide on a society with a high suicide rate: epidemiological findings on copycat suicides from South Korea. *Journal of Affective Disorders*, *156*, 56–61.

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87.

Kaasa, A., Vadi, M., & Varblane, U. (2014). Regional Cultural Differences Within European Countries: Evidence from Multi-Country Surveys. *Management International Review*, *54*(6), 825–852.

Kalish, R., & Kimmel, M. (2010). Suicide by mass murder: Masculinity, aggrieved entitlement, and rampage school shootings. *Health Sociology Review: The Journal of the Health Section of the Australian Sociological Association*, *19*(4), 451–464.

Kaplan, M. S., Huguet, N., Caetano, R., Giesbrecht, N., Kerr, W. C., & McFarland, B. H. (2016). Heavy Alcohol Use Among Suicide Decedents Relative to a Nonsuicide Comparison Group: Gender-Specific Effects of Economic Contraction. *Alcoholism, Clinical and Experimental Research*, *40*(7), 1501–1506.

Kaufman, J. A., Salas-Hernández, L. K., Komro, K. A., & Livingston, M. D. (2020). Effects of increased minimum wages by unemployment rate on suicide in the USA. *Journal of Epidemiology and Community Health*. https://doi.org/10.1136/jech-2019-212981

Kleck, G. (2018). The effect of firearms on suicide. In *Gun Studies* (pp. 309–329). Routledge.

Knapp, E. A., Bilal, U., Dean, L. T., Lazo, M., & Celentano, D. D. (2019). Economic Insecurity and Deaths of Despair in US Counties. *American Journal of Epidemiology*, *188*(12), 2131–2139.

Knopov, A., Sherman, R. J., Raifman, J. R., Larson, E., & Siegel, M. B. (2019). Household Gun Ownership and Youth Suicide Rates at the State Level, 2005–2015. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2018.10.027

Kposowa, A., Hamilton, D., & Wang, K. (2016). Impact of Firearm Availability and Gun Regulation on State Suicide Rates. *Suicide & Life-Threatening Behavior*, *46*(6), 678–696.

Kposowa, A. J. (2000). Marital status and suicide in the National Longitudinal Mortality Study. *Journal of Epidemiology and Community Health*, *54*(4), 254–261.

Kposowa, A. J. (2013). Association of suicide rates, gun ownership, conservatism and individual suicide risk. *Social Psychiatry and Psychiatric Epidemiology*, *48*(9), 1467–1479.

Kposowa, A. J., Aly Ezzat, D., & Breault, K. (2019). New Findings On Gender: The Effects Of Employment Status On Suicide. *International Journal of Women's Health*, *11*, 569–575.

Kposowa, A. J., Ezzat, D. A., & Breault, K. D. (2020). Marital status, sex, and suicide: new longitudinal findings and Durkheim's marital status propositions. *Sociological Spectrum: The Official Journal of the Mid-South Sociological Association*, *40*(2), 81–98.

Kuehn, B. M. (2011). Patients go online seeking support, practical advice on health conditions. *JAMA: The Journal of the American Medical Association*, *305*(16), 1644–1645.

Kwan, M.-P., Peterson, R. D., Browning, C. R., Burrington, L. A., Calder, C. A., & Krivo, L. J. (2008). Reconceptualizing Sociogeographic Context for the Study of Drug Use, Abuse, and Addiction. In Y. F. Thomas, D. Richardson, & I. Cheung (Eds.), *Geography and Drug Addiction* (pp. 437–446). Springer Netherlands.

Lang, M. (2013). Firearm Background Checks and Suicide. *The Economic Journal*, *123*(573), 1085–1099.

Lee, J.-Y. (2020). Search trends preceding increases in suicide: A cross-correlation study of monthly Google search volume and suicide rate using transfer function models. *Journal of Affective Disorders*, *262*, 155–164.

Lee, J.-Y., Pham, A., Wong, J., & Deng, Z. (2021). Firearm-related internet searches as a correlate of future firearm suicides: Cross-correlation analyses of monthly Google search volumes and method-specific suicide rates in the United States. *Journal of Affective Disorders Reports*, *5*, 100181.

Lester, D., & Gunn, J. F. (2016). Sociology of suicide. In *Stress: Concepts, Cognition, Emotion, and Behavior* (pp. 273–276). Elsevier.

Lewin. (1948). Resolving social conflicts: field theory in social science. Washington DC. *: American Psychological Association, first Published in*.

Li, Q., Li, R., Ji, K., & Dai, W. (2015). Kalman Filter and Its Application. *2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, 74–77.

Majority of Adults Look Online for Health Information. (2013, February 1). Pew Research Center. https://www.pewresearch.org/fact-tank/2013/02/01/majority-of-adults-look-online-for-health-information/

Mann, J. J., & Michel, C. A. (2016). Prevention of Firearm Suicide in the United States: What Works and What Is Possible. *The American Journal of Psychiatry*, *173*(10), 969–979.

Marrow, A. J. (1977). *The practical theorist: The life and work of Kurt Lewin*. *290*. https://psycnet.apa.org/fulltext/1978-31671-000.pdf

Martin, J. L. (2003). What Is Field Theory? *The American Journal of Sociology*, *109*(1), 1–49.

Mavragani, A., & Gkillas, K. (2020). COVID-19 predictability in the United States using Google Trends time series. *Scientific Reports*, *10*(1), 20693.

McCarten, J. M., Hoffmire, C. A., & Bossarte, R. M. (2015). Changes in overall and firearm veteran suicide rates by gender, 2001-2010. *American Journal of Preventive Medicine*, *48*(3), 360–364.

Miller, M., Azrael, D., & Barber, C. (2012). Suicide mortality in the United States: the importance of attending to method in understanding population-level disparities in the burden of suicide. *Annual Review of Public Health*, *33*, 393–408.

Miller, M., Hepburn, L., & Azrael, D. (2017). Firearm Acquisition Without Background Checks: Results of a National Survey. *Annals of Internal Medicine*, *166*(4), 233–239.

Miller, M., Lippmann, S. J., Azrael, D., & Hemenway, D. (2007). Household firearm ownership and rates of suicide across the 50 United States. *The Journal of Trauma*, *62*(4), 1029–1034; discussion 1034–1035.

Miller, P. G., & Sønderlund, A. L. (2010). Using the internet to research hidden populations of illicit drug users: a review. *Addiction* , *105*(9), 1557–1567.

Mohr, J. W., Bail, C. A., Frye, M., Lena, J. C., Lizardo, O., McDONNELL, T. E., Mische, A., Tavory, I., & Wherry, F. F. (2020a). Measuring Culture in Objects. In *Measuring Culture* (pp. 58–93). Columbia University Press.

Mohr, J. W., Bail, C. A., Frye, M., Lena, J. C., Lizardo, O., McDONNELL, T. E., Mische, A., Tavory, I., & Wherry, F. F. (2020b). Measuring Culture in People. In *Measuring Culture* (pp. 21–57). Columbia University Press.

Mohr, J. W., Bail, C. A., Frye, M., Lena, J. C., Lizardo, O., McDONNELL, T. E., Mische, A., Tavory, I., & Wherry, F. F. (2020c). Measuring Culture in Social Relationships. In *Measuring Culture* (pp. 94–127). Columbia University Press.

Mohr, J. W., Bail, C. A., Frye, M., Lena, J. C., Lizardo, O., McDONNELL, T. E., Mische, A., Tavory, I., & Wherry, F. F. (2020d). Pivots and Choices in the Process of Research. In *Measuring Culture* (pp. 128–156). Columbia University Press.

Mohr, J. W., Bail, C. A., Frye, M., Lena, J. C., Lizardo, O., McDONNELL, T. E., Mische, A., Tavory, I., & Wherry, F. F. (2020e). The Future of Measuring Culture. In *Measuring Culture* (pp. 157–174). Columbia University Press.

Mohr, J. W., Bail, C. A., Frye, M., Lena, J. C., Lizardo, O., McDONNELL, T. E., Mische, A., Tavory, I., & Wherry, F. F. (2020f). Why Measure Culture? In *Measuring Culture* (pp. 1–20). Columbia University Press.

Mok, K., Ross, A. M., Jorm, A. F., & Pirkis, J. (2016). An Analysis of the Content and Availability of Information on Suicide Methods Online. *Journal of Consumer Health on the Internet*, *20*(1-2), 41–51.

Monnat, S. M., Peters, D. J., Berg, M. T., & Hochstetler, A. (2019). Using Census Data to Understand County-Level Differences in Overall Drug Mortality and Opioid-Related Mortality by Opioid Type. *American Journal of Public Health*, *109*(8), 1084–1091.

Montez, J. K., Zajacova, A., Hayward, M. D., Woolf, S. H., Chapman, D., & Beckfield, J. (2019). Educational Disparities in Adult Mortality Across U.S. States: How Do They Differ, and Have They Changed Since the Mid-1980s? *Demography*, *56*(2), 621–644.

Moore, J., & Abetz, J. S. (2019). What Do Parents Regret About Having Children? Communicating Regrets Online. *Journal of Family Issues*, *40*(3), 390–412.

Mukherjee, S., Becker, N., Weeks, W., & Ferres, J. L. (2020). Using internet search trends to forecast short term drug overdose deaths: A case study on Connecticut. *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1332–1339.

Nagelhout, G. E., Hummel, K., de Goeij, M. C. M., de Vries, H., Kaner, E., & Lemmens, P. (2017). How economic recessions and unemployment affect illegal drug use: A systematic realist literature review. *The International Journal on Drug Policy*, *44*, 69–83.

National Institute on Drug Abuse. (2022, January 20). *Overdose Death Rates*. National Institute on Drug Abuse. https://nida.nih.gov/drug-topics/trends-statistics/overdose-death-rates

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews. Computational Statistics*, *4*(2), 199–203.

Neumayer, E. (2003). Are Socioeconomic Factors Valid Determinants of Suicide? Controlling for National Cultures of Suicide with Fixed-Effects Estimation. *Cross-Cultural Research: Official Journal of the Society for Cross-Cultural Research / Sponsored by the Human Relations Area Files, Inc*, *37*(3), 307–329.

Nichter, B., Norman, S., Haller, M., & Pietrzak, R. H. (2019). Psychological burden of PTSD, depression, and their comorbidity in the U.S. veteran population: Suicidality, functioning, and service utilization. *Journal of Affective Disorders*, *256*, 633–640.

Nordentoft, M., Mortensen, P. B., & Pedersen, C. B. (2011). Absolute risk of suicide after first hospital contact in mental disorder. *Archives of General Psychiatry*, *68*(10), 1058–1064.

Norström, T. (1995). The Impact of Alcohol, Divorce, and Unemployment on Suicide: A Multilevel Analysis. *Social Forces; a Scientific Medium of Social Study and Interpretation*, *74*(1), 293–314.

Nosrati, E., Kang-Brown, J., Ash, M., McKee, M., Marmot, M., & King, L. P. (2019). Economic decline, incarceration, and mortality from drug use disorders in the USA between 1983 and 2014: an observational analysis. *The Lancet. Public Health*, *4*(7), e326–e333.

Owens, D., Horrocks, J., & House, A. (2002). Fatal and non-fatal repetition of self-harm. Systematic review. *The British Journal of Psychiatry: The Journal of Mental Science*, *181*, 193–199.

Parker, J., Cuthbertson, C., Loveridge, S., Skidmore, M., & Dyar, W. (2017). Forecasting state-level premature deaths from alcohol, drugs, and suicides using Google Trends data. *Journal of Affective Disorders*, *213*, 9–15.

Paul, E., Mergl, R., & Hegerl, U. (2017). Has information on suicide methods provided via the Internet negatively impacted suicide rates? *PloS One*, *12*(12), e0190136.

Phillips, M. R., Shen, Q., Liu, X., Pritzker, S., Streiner, D., Conner, K., & Yang, G. (2007). Assessing depressive symptoms in persons who die of suicide in mainland China. *Journal of Affective Disorders*, *98*(1-2), 73–82.

Popovici, I., Maclean, J. C., Hijazi, B., & Radakrishnan, S. (2018). The effect of state laws designed to prevent nonmedical prescription opioid use on overdose deaths and treatment. *Health Economics*, *27*(2), 294–305.

Prado-Román, C., Gómez-Martínez, R., & Orden-Cruz, C. (2021). Google Trends as a Predictor of Presidential Elections: The United States Versus Canada. *The American Behavioral Scientist*, *65*(4), 666–680.

Preventing Suicide |Violence Prevention|Injury Center|CDC. (2020, April 21). https://www.cdc.gov/violenceprevention/suicide/fastfact.html

Quinones, S. (2015). *Dreamland: The True Tale of America's Opiate Epidemic*. Bloomsbury Publishing USA.

Racine, M. (2018). Chronic pain and suicide risk: A comprehensive review. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *87*(Pt B), 269–280.

Rees, D. I., Sabia, J. J., Argys, L. M., Dave, D., & Latshaw, J. (2019). With a Little Help from My Friends: The Effects of Good Samaritan and Naloxone Access Laws on Opioid-Related Deaths. *The Journal of Law and Economics*, *62*(1), 1–27.

Richard Fry, S. W. (2010). *Women, Men and the New Economics of Marriage*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.697.463

Rossen, L. M., Hedegaard, H., Khan, D., & Warner, M. (2018). County-Level Trends in Suicide Rates in the U.S., 2005-2015. *American Journal of Preventive Medicine*, *55*(1), 72–79.

Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science*, *308*(5722), 648–652.

Sarkisian, N., & Gerstel, N. (2016). Does singlehood isolate or integrate? Examining the link between marital status and ties to kin, friends, and neighbors. *Journal of Social and Personal Relationships*, *33*(3), 361–384.

Song, T. M., Song, J., An, J.-Y., Hayman, L. L., & Woo, J.-M. (2014). Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei Medical Journal*, *55*(1), 254–263.

Springer, K. W., Lee, C., & Carr, D. (2019). Spousal Breadwinning Across 30 Years of Marriage and Husbands' Health: A Gendered Life Course Stress Approach. *Journal of Aging and Health*, *31*(1), 37–66.

Stack, S., & Kposowa, A. J. (2016). Culture and Suicide Acceptability: A Cross-National, Multilevel Analysis. *The Sociological Quarterly*, *57*(2), 282–303.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, *118*, 26–40.

Stephens-Davidowitz, S., & Pinker, S. (2017). *Everybody lies: big data, new data, and what the internet can tell Us about who we really are*. HarperCollins New York.

Stickley, A., Koyanagi, A., Ueda, M., Inoue, Y., Waldman, K., & Oh, H. (2020). Physical multimorbidity and suicidal behavior in the general population in the United States. *Journal of Affective Disorders*, *260*, 604–609.

Sueki, H. (2012). Association between deliberate self-harm-related Internet searches and the mental states and lifetime suicidal behaviors of Japanese young adults. *Psychiatry and Clinical Neurosciences*, *66*(5), 451–453.

Suicide. (n.d.-a). National Institute of Mental Health (NIMH). Retrieved July 27, 2022, from https://www.nimh.nih.gov/health/statistics/suicide

Suicide, S. S. (n.d.-b). *Veteran Suicide Stats*. Stop Soldier Suicide. Retrieved July 28, 2022, from https://stopsoldiersuicide.org/vet-stats

Taylor, C. J. (2014). Physiological stress response to loss of social influence and threats to masculinity. *Social Science & Medicine*, *103*, 51–59.

Thornton, L., Handley, T., Kay-Lambkin, F., & Baker, A. (2017). Is A Person Thinking About Suicide Likely to Find Help on the Internet? An Evaluation of Google Search Results. *Suicide & Life-Threatening Behavior*, *47*(1), 48–53.

Topic: Internet usage in the United States. (n.d.). Www.statista.com. Retrieved December 16, 2019, from https://www.statista.com/topics/2237/internet-usage-in-the-united-states/

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking Internal Consistency in Cronbach's Alpha. *Leisure Sciences*, *39*(2), 163–173.

Vitt, D. C., McQuoid, A. F., Moore, C., & Sawyer, S. (2018). Trigger warning: the causal impact of gun ownership on suicide. *Applied Economics*, *50*(53), 5747–5765.

Vogel, D. L., Heimerdinger-Edwards, S. R., Hammer, J. H., & Hubbard, A. (2011). "Boys don't cry": Examination of the links between endorsement of masculine norms, self-stigma, and help-seeking attitudes for men from diverse backgrounds. *Journal of Counseling Psychology*, *58*(3), 368–382.

Wasserman, I. M. (1984). The influence of economic business cycles on United States suicide rates. *Suicide & Life-Threatening Behavior*, *14*(3), 143–156.

Wiebe, D. J. (2003). Homicide and suicide risks associated with firearms in the home: a national case-control study. *Annals of Emergency Medicine*, *41*(6), 771–782.

Wilson, W. J. (2011). *When Work Disappears: The World of the New Urban Poor*. Knopf Doubleday Publishing Group.

WISQARS (Web-based Injury Statistics Query and Reporting System). (2020, July 1). https://www.cdc.gov/injury/wisqars/index.html

Wolfe-Clark, A. L., & Bryan, C. J. (2017). Integrating Two Theoretical Models to Understand and Prevent Military and Veteran Suicide. *Armed Forces and Society*, *43*(3), 478–499.

Wood, D. S., Wood, B. M., Watson, A., Sheffield, D., & Hauter, H. (2020). Veteran Suicide Risk Factors: A National Sample of Nonveteran and Veteran Men Who Died by Suicide. *Health & Social Work*, *45*(1), 23–30.

Wray, M., Colen, C., & Pescosolido, B. (2011). The Sociology of Suicide. *Annual Review of Sociology*, *37*(1), 505–528.

Yeung, T. Y.-C. (2019). Measuring Christian Religiosity by Google Trends. *Review of Religious Research*, *61*(3), 235–257.

Zamarripa, M. X., Wampold, B. E., & Gregory, E. (2003). Male gender role conflict, depression, and anxiety: Clarification and generalizability to women. *Journal of Counseling Psychology*, *50*(3), 333–338.