Absolutist Words From Search Volume Data Predict State-Level Suicide Rates in the United States

2 authors:

Jaïs Adam-Troïan
Keele University

60 PUBLICATIONS **196** CITATIONS

SEE PROFILE

Thomas Arciszewski
Aix-Marseille Université

48 PUBLICATIONS **316** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Disengagement from violent extremism / deradicalization View project

Early Social Cognition View project

Absolutist Words from Search Volume Data Predict State-Level Suicide Rates in the US

Jais Adam-Troian[1], Thomas Arciszewski[2]

Abstract

Suicide continues to be a major public health issue, especially in the US. It is a well- established fact that depression and suicidal ideation are risk factors for suicide. Drawing upon recent research which shows that absolutist words (e.g. '*completely*', '*totally*') constitute linguistic markers of suicidal ideation, we created an online index of absolutist thinking (ATI) using search query data (i.e. Google Trends time series). Mixed-model analyses of age-adjusted suicide rates in the US from 2004-2017, revealed that ATI is linked with suicides β = .22, 95%CI[.12, .31], *p* < .001 and predicts suicides within one year β = .16, 95%CI[.05, .28], *p* = .006 independently of state characteristics and historical trends. It is the first time that a collective measure of absolutist thinking is used to predict real-world suicide outcomes. Therefore, the present study paves the way for novel research avenues in clinical psychological research.

Keywords: suicide, absolutist thinking, prediction, search query, US states.

Corresponding Author*:* Jais Adam-Troian, American University of Sharjah, Sharjah, UAE jadam@aus.edu, +905539054791.

Conflict of Interests: The authors have no conflict of interest to declare.

---

[1] European Institute, Istanbul Bilgi University, Turkey
[2] PSYCLE Center, Aix-Marseille University, France

Introduction

Suicide remains a significant public health issue, with an estimated 40,000 yearly casualties in the United States alone, leading to an approximate cost of over $44 billion each year for the US economy due to medical and work loss costs (CDC, 2017). Therefore, finding ways to monitor, anticipate and predict suicide deaths is of central interest to engage in preventive actions aimed at the general population.

One current approach to do so consists in leveraging information from social media data to compute online indexes that are linked with suicide rates (Won, Myung, Song, Lee, Kim, Carroll, & Kim, 2013). Though this approach yields interesting results, it remains complicated to use for applied purposes given privacy concerns regarding social media. Moreover, there is heterogeneity in what constitutes appropriate indexes, which are often more empirically (data) driven than based on reliable psychological theorizing. For instance, Won et al.'s (2013) suicide indicator (computed from suicide-related content on social media posts) was very susceptible to celebrity suicide events and displayed high variability, indicating the broad and noisy character of this measure.

This is why we decided to test whether a simple, yet robust online index could be computed from readily accessible online data, which would predict real-life suicide deaths. To achieve this, we turned to clinical psychological theory to identify potent constructs that can be tapped into by linguistic markers. Among several identified risk factors for suicide deaths some of the most potent remain depression and suicide ideation (Ribeiro, Huang, Fox, & Franklin, 2018). Incidentally, a promising candidate for operationalizing linguistic traces of both is absolutist thinking (Beck, 1979). In a nutshell, absolutist thoughts are context-independent and devoid of nuance. Absolutist and non-absolutist concepts both refer to magnitudes or probabilities, but absolutist thoughts differ from non-absolutist ones in that they preclude any kind of gradation ('completely' *vs.* 'somewhat' for instance).

Currently, it is established that increased absolutist thinking is associated with anxiety, depression and suicidal ideation (Al-Mosaiwi & Johnstone, 2018). In line with this, absolutist words have been found more prevalent among suicidal patients (Neuringer, 1964) and in writings of suicidal individuals (Wedding, 2000). Though the exact cognitive mechanisms at play remain the target of ongoing investigations, cognitive rigidity and impairment of problem-solving abilities are co-morbid with depression and suicidal ideation (Ellis & Rutherford, 2008; Williams & Pollock, 2008). Consequently, recent research has validated a dictionary of 19 words which are linguistic markers of absolutist thinking (e.g. '*completely*', '*totally*') and therefore more prevalent among online posts from forums about depression and suicide (Al-Mosaiwi & Johnstone, 2018). Thus, it is theoretically expected that an aggregated index using these words' prevalence from online data sources could be a useful proxy for collective-level depression and suicidal ideation.

The present research

To compute thus indicator, we decided to use search volume data from Google Trends. In short, Google Trends provides researchers with the frequency at which a specific search term is typed, compared to all different search terms across specified languages and geographical areas (Google, 2017). These data are publicly available and can be easily retrieved, making them a perfect fit for our purpose. Furthermore, search entries are predictive of real-life phenomena, such as stock markets fluctuations (from words like '*crisis*' or '*subprime*' see Preis, Reith, & Stanley, 2010) and consumer behaviour (Jun, Yoo, & Choi, 2018). Also, when extracting data regarding a specific word in Google Trends it is possible to retrieve its prevalence among *all search entries* in a given population. For example, data extracted for the word '*completely*' would count each time this word was contained in every request among the population and area of interest during a specified time span (for more information regarding the

inner workings and applications of Google Trends data see also Yun et al., 2018; Troian, Arciszewski, & Apostolidis, 2019).

Thus, we decided to investigate whether aggregated search volume data computed from the absolutist word dictionary could predict suicide deaths at the state level across US. In line with Al-Mosaiwi & Johnstone (2018), we hypothesized that an aggregated linguistic measure of state-level absolutist thinking should be positively linked with suicide rates, and that this measure should also be predictive of state-level suicide rates in time.

METHODS

Data sources.

We collected search volume time series for word searches regarding the 19 terms from the absolutist language dictionary (available in Al-Mosaiwi, & Johnstone, 2018; see appendix 1) with the help of Google Trends. Here, it must be noted that we aimed to create an index counting the occurrences of absolutist words in the written language of search engine users, independent of their emotional valence. Indeed, specifying emotion-related words for each of the 19 absolutist words would create methodological issues, such as a drastic loss in data points (due to decreased likelihood of users typing specific combination of words) and the need to take into account a whole spectrum of emotions (i.e. not simply 'negative' and 'positive') for accuracy purposes. Given that the use of absolutist language is an indicator of depression/suicidal ideation in and of itself regardless of the semantic context in which absolutist words occur, we opted for a more parsimonious analysis relying solely on the 19 terms from the absolutist language dictionary.

Whenever data was too few and the output indicated an estimate '< *1*', we rounded it down to 0 to allow for quantitative analysis (those occurrences were overall rare though, < 1% of total data). The extracted data thus consisted in 19 time series from which we computed a

single sum score variable (Absolutist Thinking Index, or ATI) after checking reliability (α = .90). The suicide rates across US states were obtained using the CDC data on fatal injuries (https://webappa.cdc.gov/sasweb/ncipc/DataRestriction_inj.html). We requested data for age adjusted suicide rates (thus preventing confounding) merged across genders and ethnic groups, split by state and year. The time period covered by all our data ranged from 2004 (first Google Trends data available) to 2017 (last suicide rate from the CDC available), yielding 14 repeated measures for each of the 50 US states (N = 700; data for the District of Columbia were not available from Google Trends).

Analytic Strategy.

We decided to conduct two separate analyses using mixed modelling to account for the clustered nature of our data (Schielzeth & Nakagawa, 2013). This analysis would also have the advantage of specifying states as a random factor, thereby adjusting our results on a host of unmeasured characteristics that could confound the observed relationship across time points (levels of economic inequality, demographic characteristics of the population, prevalence of substance abuse…). The first analysis would be performed on the data as such, with states as a random factor, time as a-within subject (state) repeated measure (covariate), ATI as a continuous predictor and suicide rates as the dependent variable. The second analysis would be *time-lagged*. Indeed, because our data is observational, one way to explore causal plausibility is to conduct lagged predictive analyses in two different models, once with the dependent variable (suicide rates) at *t*+1 from the independent, and once more by testing the effect of the previous dependent variable (suicide rates) on the independent (ATI, now dependent) at *t*+1. Then, by comparing the two models, one can obtain some evidence to support or refute the hypothesized causal direction (depending on what model fits the data best).

RESULTS

Association test.

ATI was *z*-transformed. Then, a mixed-model was computed (Maximum Likelihood, AIC = 2369.35, Log-Likelihood = 1179.2, $r^2_{conditional}$ = .93, $r^2_{marginal}$ = .10) according to the following: *suicide ~ 1 + (1 | state) + ATI + time*. It yielded a significant main effect of time $t(650) = 24.09$, $p < .001$, β = .29, 95%CI[.27,.32], *SE* = .01 (reflecting the observed increase in US suicide rates over the past decade; CDC, 2018) and the predicted positive effect of ATI emerged, $t(650) = 4.31$, $p < .001$, β = .22, 95%CI[.12,.31], *SE* = .05. As can be seen on figure 1, this association is positive in every State.

These results highlighted a robust positive association between ATI and suicide rates, even when adjusting for between-state structural differences and historical trends (increase with time). Furthermore, we ran our analyses once more to check the robustness of our findings when adjusting for known environmental predictors of suicide deaths. Even when including GSP per capita and firearms ownership rates, the positive effect of ATI remained, $t(650) = 4.26$, $p < .001$, β = .21, 95%CI[.12,.31], *SE* = .05 (see appendix 2).
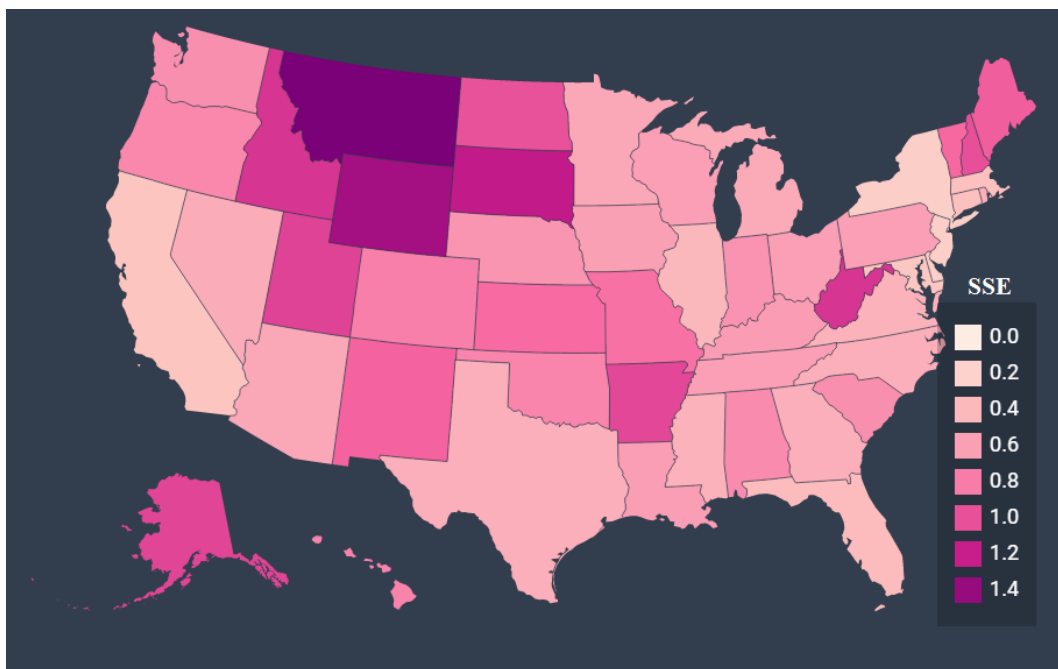

Figure 1. *Map displaying the association between ATI and suicide rates across US States. SSE = Standardized Slope Estimates.*

Time-lagged analyses.

To assess the plausibility of a causal direction, we then computed two lagged variables from ATI and suicide rates (1-year lag). Model 1 was computed according to the following: *suicide ~ 1 + (1 | state) + lagged_ATI + time* and Model 2: *ATI ~ 1 + (1 | state) + lagged_suicide + time*. As reported in table 1, ATI was a better predictor of suicide rates $t(604) = 2.78$, $p = .006$, $\beta = .16$, 95%CI[.05, .28], $SE = .06$, than suicide rates of ATI $t(650) = 2.44$, $p = .015$, $\beta = -.09$, 95%CI[-.16, -.02], $SE = .04$, which happened to be negative (it may be that increased suicide rates lowers ATI the following year because of a drop in number of depressed individuals in the population due to suicide). Also, Model 1 was more parsimonious than Model 2 ($AIC_{M1} < AIC_{M2}$), while analyses indicated overfitting issues with Model 2 (hence the $r^2_{conditional} = r^2_{marginal}$). Therefore, this second analysis provided evidence for considering a potentially causal role of ATI upon suicide rate increases, in line with current psychopathological models of suicide behaviour.

Table 1. *Mixed models for time-lagged analyses(Outcomes: Model 1 = Lagged Suicide rates; Model 2 = Lagged ATI).*

	t(Df)	AIC	LL	B(SE)	95%CI	$r^2_{conditional}$	$r^2_{marginal}$	p
Model 1		2208.3	-1099.2			.93	.10	
ATI	2.78(604)			.16(.06)	[.05, .28]			.006
Time	25.4(601)			.32(.01)	[.30, .35]			< .001
Model 2		9480	-4735			.26	.26	
SR	2.44(650)			-.09(.04)	[-.16, -.02]			.015
Time	15(650)			.14(.01)	[.12, .16]			< .001

Note. LL = Log-Likelihood, SR = Suicide Rates, β = standardized regression coefficients, SE = Standard Error.

DISCUSSION

The present research highlights for the first time the feasibility of predicting suicide rates at the collective level with the help of search volume indices (Google Trends) drawing

upon theoretically driven terms (ATI). Also, absolutist thinking is characteristic of depression and suicidal ideation. It is thus the first time collective-level linguistic traces of specific cognitive markers for risk-factors of suicide behaviour are used for predictive purposes at the population level. Readily accessible search data (i.e. natural linguistic traces) also constitutes an ecological measure less susceptible to biases commonly encountered with self-report methods (e.g. social desirability). As the results showed, our state-level absolutist-thinking indicator seems predictive of suicide rates across the US. Because our variables were z-transformed, it is even possible to quantify this phenomenon. Given the results from the first analysis, an increase of one standard deviation in ATI is directly *associated* with 22% more suicides. As can be seen from Model 1 in our second analysis, this same increase in ATI *leads* to 16% more suicide on average the next year, in any given state.

These findings open the way for diagnosing high-risk geographical areas and using ATI to assess the efficacy of public policies targeting mental health and the impact of societally traumatic events (terror attacks, economic crises and so on). Indeed, most of current research on depression and suicide remains focused on individual-level measures and outcomes. While individual diagnosis and treatment of depression is useful for reducing suicide deaths (such as cognitive-behavioural therapy see Brown, Ten Have, Henriques, Xie, Hollander, & Beck, 2005), it is also known that other, social factors can impact depression and suicide deaths. For instance, social isolation and loneliness have been found to cause mortality by suicide among older adults (Steptoe, Shankar, Demakakos, Wardle, 2013). Such social risk factors can be lowered by interventions at the collective level, including to combat suicide deaths due to social isolation (Dickens, Richards, Greaves, Campbell, 2011; see also Haslam, Jetten, Cruwys, Dingle, & Haslam, 2018).

If suicide deaths are now better understood scientifically and if efficient methods for preventing them have been developed, collective level indicators of depression and suicidal

ideation could be useful to monitor and eventually assess the impact of collective level interventions and prevention campaigns. Because the ATI is cheaper and relatively cost-effective than large annual nationally representative surveys, routinely updated (weekly), constantly collected and publicly available, it could enable many novel possibilities.

For instance, research suggests that suicide prevention strategies can be effective, yet much less so regarding media and public awareness campaigns (due to lack of data and especially RCT data; see Zalsman et al., 2016). Because of its availability since 2004, the ATI might constitute a useful indicator to conduct archival analysis of previously conducted prevention campaigns, down to the metro level. Classifying different types of campaigns based on their content and assessing their effects on ATI might yield useful knowledge as to what makes up an efficient prevention message, even allowing for factoring in cultural and environmental factors likely to moderate campaign efficacy.

In addition, this approach is not limited to suicide prevention, since ATI might tap into anxiety disorders at large and may thus constitute a broad measure of population mental health. Using indicators of the ATI type might help decision makers identify at risk-areas, design targeted community interventions to optimize mental healthcare resources allocation. Moreover, identifying and comparing communities in which ATI remain stable (*vs.* increase) in the face of unexpected events known to affect mental health (e.g. natural disasters, Neria & Shultz, 2012) could help pinpoint environmental facilitators of community resilience and inform mental healthcare policymaking itself. To this end however, much more research is needed to quantify the link between ATI and population prevalence of mental illnesses.

Limitations

Despite these encouraging first results, several concerns must be acknowledged. First, absolutist thinking is more prevalent among individuals with depression and suicidal ideation,

but also among individuals with bipolar and eating disorders (Palascha, van Kleef, & van Trijp, 2015; Alberts. Thewissen, & Raes, 2012). It is furthermore associated with negative emotions in general (Cohen, 2012), indicating that this measure might not be highly specific. In addition, our data was aggregated yearly, due to data accessibility constraints. But because research shows dramatic variations in severity and type of suicide ideation over very short periods of time (e.g. daily, see Kleiman, Turner, Fedor, Beale, Huffman, & Nock, 2017), further investigations using weekly or monthly collected data points should be conducted to get a finer grained understanding of the dynamics at play between absolutist thinking and suicide behaviour at the collective level.

Notwithstanding the observational nature of our investigation, which prevent any causal interpretation of the findings, the US context is also particular, with an ongoing 'opioid epidemic' and increasing social inequalities (which are known to generate suicide deaths, see Pickett & Wilkinson, 2015). Thus, cross-cultural replications of our analyses should be conducted to ensure the generalizability of the present findings, using other kinds of 'big data' indicators to investigate the weight of methodological and cultural factors.

Conclusion

For the first time, we provide evidence of a link between absolutist thinking - a marker of depression/suicidal ideation - and suicide deaths at the collective level. Within the boundaries of the above-mentioned limitations, we remain confident that the present findings open a novel avenue for psychological research and public health interventions regarding the prevention of suicide deaths.

REFERENCES

Alberts, H. J., Thewissen, R., & Raes, L. (2012). Dealing with problematic eating behaviour.

The effects of a mindfulness-based intervention on eating behaviour, food cravings, dichotomous thinking and body image concern. *Appetite*, *58*(3), 847-851.

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, *6*(4), 529-542.

Beck, A. T. (1979). Cognitive therapy and the emotional disorders. London, England: Penguin.

Brown, G. K., Ten Have, T., Henriques, G. R., Xie, S. X., Hollander, J. E., & Beck, A. T. (2005). Cognitive therapy for the prevention of suicide attempts: a randomized controlled trial. *Jama*, *294*(5), 563-570.

Center for Disease Control and Prevention (2017). *Suicide and self-inflicted injury: Fast Stats.* Retrieved from https:// www.cdc.gov/nchs/fastats/suicide.htm

Center for Disease Control and Prevention (2018). *Suicide Prevention*. Retrieved from https://www.cdc.gov/media/releases/2018/p0607-suicide-prevention.html

Cohen, S. J. (2012). Construction and preliminary validation of a dictionary for cognitive rigidity: Linguistic markers of overconfidence and overgeneralization and their concomitant psychological distress. *Journal of psycholinguistic research*, *41*(5), 347-370.

Dickens, A. P., Richards, S. H., Greaves, C. J., & Campbell, J. L. (2011). Interventions targeting social isolation in older people: a systematic review. *BMC public health*, *11*(1), 647.

Ellis, T. E., & Rutherford, B. (2008). Cognition and suicide: Two decades of progress. *International Journal of Cognitive Therapy*, *1*(1), 47-68.

Google (2017). *Google Trends Help*. Retrieved at

https://support.google.com/trends/?hl=ko#topic=6248052

Haslam, C., Jetten, J., Cruwys, T., Dingle, G., & Haslam, S. A. (2018). *The new psychology of*

health: Unlocking the social cure. London: Routledge.

Hawton, K., & Van Heeringen, K. (Eds.). (2000). *The international handbook of suicide and*

attempted suicide. John Wiley & Sons.

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.

Kleiman, E. M., Turner, B. J., Fedor, S., Beale, E. E., Huffman, J. C., & Nock, M. K. (2017).

Examination of real-time fluctuations in suicidal ideation and its risk factors: Results

from two ecological momentary assessment studies. *Journal of Abnormal*

Psychology, *126*(6), 726.

Neuringer, C. (1964). Rigid thinking in suicidal individuals. *Journal of Consulting*

Psychology, *28*(1), 54.

Neria, Y., & Shultz, J. M. (2012). Mental health effects of Hurricane Sandy: characteristics,

potential aftermath, and response. *Jama*, *308*(24), 2571-2572.

Palascha, A., van Kleef, E., & van Trijp, H. C. (2015). How does thinking in Black and White

terms relate to eating behavior and weight regain?. *Journal of health psychology*, *20*(5),

638-648.

Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on

different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1933), 5707-5719.

Schielzeth, H., & Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, *4*(1), 14-24.

Steptoe, A., Shankar, A., Demakakos, P., & Wardle, J. (2013). Social isolation, loneliness, and all-cause mortality in older men and women. *Proceedings of the National Academy of Sciences*, *110*(15), 5797-5801.

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: a causal review. *Social science & medicine*, *128*, 316-326.

Ribeiro, J. D., Huang, X., Fox, K. R., & Franklin, J. C. (2018). Depression and hopelessness as risk factors for suicide ideation, attempts and death: meta-analysis of longitudinal studies. *The British Journal of Psychiatry*, *212*(5), 279-286.

Troian, J., Arciszewski, T., & Apostolidis, T. (2019). The dynamics of public opinion following terror attacks: Evidence for a decrease in equalitarian values from Internet Search Volume Indices. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *13*(3).

Wedding, D. (2000). Cognitive distortions in the poetry of Anne Sexton. *Suicide and Life-Threatening Behavior*, *30*(2), 140-144.

Won, H. H., Myung, W., Song, G. Y., Lee, W. H., Kim, J. W., Carroll, B. J., & Kim, D. K.

(2013). Predicting national suicide numbers with social media data. *PloS one*, *8*(4),

e61809.

Zalsman, G., Hawton, K., Wasserman, D., van Heeringen, K., Arensman, E., Sarchiapone, M.,

... & Purebl, G. (2016). Suicide prevention strategies revisited: 10-year systematic

review. *The Lancet Psychiatry*, *3*(7), 646-659.