County-level estimates of suicide mortality in the USA: a modelling study

Sasikiran Kandula, MS,
Department of Environmental Health Sciences, Columbia University, New York, NY, USA

Gonzalo Martinez-Alés, PhD,
Department of Epidemiology, Columbia University, New York, NY, USA; CAUSALab, Harvard T H Chan School of Public Health, Boston, MA, USA; Mental Health Network Biomedical Research Center, Madrid, Spain; Mental Health Research Group, Hospital La Paz Institute for Health Research, Madrid, Spain

Caroline Rutherford, MS,
Department of Epidemiology, Columbia University, New York, NY, USA

Catherine Gimbrone, MPH,
Department of Epidemiology, Columbia University, New York, NY, USA

Mark Olfson, MD,
Department of Epidemiology, Columbia University, New York, NY, USA; Department of Psychiatry, Columbia University, New York, NY, USA

Madelyn S Gould, PhD,
Department of Epidemiology; Department of Psychiatry, Columbia University, New York, NY, USA

Katherine M Keyes, PhD,
Department of Epidemiology, Columbia University, New York, NY, USA

Jeffrey Shaman, PhD
Department of Environmental Health Sciences, Columbia University, New York, NY, USA

Summary

Background—Suicide is one of the leading causes of death in the USA and population risk prediction models can inform decisions on the type, location, and timing of public health

interventions. We aimed to develop a prediction model to estimate county-level suicide risk in the USA using population characteristics.

Methods—We obtained data on all deaths by suicide reported to the National Vital Statistics System between Jan 1, 2005, and Dec 31, 2019, and age, sex, race, and county of residence of the decedents were extracted to calculate baseline risk. We also obtained county-level annual measures of socioeconomic predictors of suicide risk (unemployment, weekly wage, poverty prevalence, median household income, and population density) and state-level prevalence of major depressive disorder and firearm ownership from US public sources. We applied conditional autoregressive models, which account for spatiotemporal autocorrelation in response and predictors, to estimate county-level suicide risk.

Findings—Estimates derived from conditional autoregressive models were more accurate than from models not adjusted for spatiotemporal autocorrelation. Inclusion of suicide risk and protective covariates further reduced errors. Suicide risk was estimated to increase with each SD increase in firearm ownership (2·8% [95% credible interval (CrI) 1·8 to 3·9]), prevalence of major depressive episode (1·0% [0·4 to 1·5]), and unemployment rate (2·8% [1·9 to 3·8]). Conversely, risk was estimated to decrease by 4·3% (−5·1 to −3·2) for each SD increase in median household income and by 4·3% (−5·8 to −2·5) for each SD increase in population density. An increase in the heterogeneity in county-specific suicide risk was also observed during the study period.

Interpretation—Area-level characteristics and the conditional autoregressive models can estimate population-level suicide risk. Availability of near real-time situational data are necessary for the translation of these models into a surveillance setting. Monitoring changes in population-level risk of suicide could help public health agencies select and deploy targeted interventions quickly.

Introduction

Suicide rates in the USA have increased by more than 30% in the past two decades, with suicide ranking among the ten most common causes of death for this period.[1] In addition to deaths due to drug overdoses and alcohol use, which have also increased substantially in the same time period, suicides have contributed to the decrease in overall life expectancy observed in the USA.[2] Reducing deaths by suicide is therefore an urgent public health challenge, and methods to predict suicide risk could be vital for determining optimal allocation of suicide prevention resources.

To date, models to predict suicide risk have largely been focused at the individual level, using patient demographic characteristics and clinical history to estimate patient risk.[3] Some of these models have been deployed operationally to screen patients,[4] with evidence suggesting their wider adoption can be hastened through improvements in predictive ability.[5,6] In contrast to individual-level models, population-level risk models have been less frequently used despite strong motivating factors in their favour, including evidence of efficacy of population-level suicide prevention interventions, such as restrictions on access to lethal means (eg, firearms).[7-9] Population-level models can complement individual-level models since they can inform decisions on the type, location, and timing of public

health interventions, and thus can be considered as valuable counterparts to clinical case management.[10,11] Additionally, when variables providing situational awareness, such as calls to crisis hotline services or posts to social media sites are also included in these models, near real-time changes in population risk can potentially be detected sooner, thus aiding the deployment of timely and responsive interventions. Similarly, geographically well-resolved risk estimates can support deployment of more targeted interventions.

Risk factors for suicide have been extensively studied and include demographic characteristics such as age, race or ethnicity, sex, socioeconomic status,[12-16] and mental health history.[17,18] However, studies assessing the effect sizes of combinations of these characteristics are relatively scarce. Meta-analyses of reported effect sizes have identified considerable heterogeneity.[19] Differences also exist in population-level association studies of suicide rates and risk factors.[20,21]

A crucial consideration when building population-level models is the presence of spatiotemporal autocorrelation in the outcome. Inadequate accounting for this phenomenon, whereby proximate areal units during close time periods are likely to have similar observations compared with those more distant in space and time, can lead to incorrect assumptions of independence and thereby to erroneous interpretations of effects.

With deaths by suicide, spatial autocorrelation can indicate an underlying spatially correlated risk factor or a form of neighbourhood effect, whereas temporal autocorrelation can be due to the same population being observed in adjacent periods, subject to the same long-term socioeconomic and environmental stressors.[22-24] Furthermore, suicidal behaviour has been described with contagion hypotheses and theories.[25] For example, acts of intentional self-harm that are directly and causally related to each other in suicide clusters,[26] increases in death by suicide immediately after sensationalistic media reporting of high-profile deaths by suicide,[27,28] or fictional depictions of suicide can be considered as a contagious process[29] and lead to spatial and temporal autocorrelation.[30]

In this study, we aimed to develop a predictive model to estimate county-level suicide risk in the USA using area-level characteristics, while accounting for spatiotemporal autocorrelation in outcome. Specifically, we aimed to use the model to quantify the effect estimates of socioeconomic status covariates on suicide risk; quantify annual national suicide risk in the USA and changes in heterogeneity of county-level risk between 2005 and 2016; and to assess the accuracy in predicting yearly county-level suicide mortality risk and measure improvements relative to common-sense baseline risk estimates.

Methods

Data sources

We used a variety of public data sources for estimates of area-level suicide mortality risk factors. Detailed mortality records were obtained through a request to the National Center for Health Statistics.

Records of all-cause deaths were obtained from the US National Vital Statistics System.[31] All deaths resulting from suicide reported between Jan 1, 2005, and Dec 31, 2019, were identified using the tenth revision of the International Classification of Diseases with underlying cause-of-death codes X60–X84, Y87.0, and U03. County estimates for total population and population stratified by age and sex were obtained from the Bridged-Race intercensal (2005–09)[32] and postcensal (2011–19)[33] datasets and used to calculate annual, county-level suicide mortality risk.

We also extracted data on explanatory variables for suicide risk (table; appendix pp 3–5). We obtained county-level estimates of the annual proportion of the population living in poverty and median household income (US$) from the Small Area Income and Poverty Estimates programme of the US Census Bureau.[34] State-level estimates of the proportion of the population with at least one major depressive episode during the previous year were obtained from the National Survey on Drug Use and Health dataset.[35] Since county-level data were not available, prevalence was assumed to be the same in all counties in a state. State-level annual estimates of the proportion of adults who live in a household with firearms were obtained from the RAND Corporation Household Firearm Ownership Database;[36] ownership rates were assumed to be the same in all counties in a state. County-level estimates of annual mean weekly wage across all industries were obtained from the Quarterly Census of Employment and Wages programme of the US Bureau of Labor.[37] We obtained estimates of county-level unemployment rates from data reported to the US Bureau of Labor Statistics.[38] Annual county-level population density was estimated from intercensal and postcensal population estimates[32,33] and county land area.

Conditional autoregressive models

We modelled area-level risk of suicides with spatiotemporal extensions of conditional autoregressive models, a family of Bayesian inference models commonly used in case of unmeasured spatial autocorrelation.[39,40] Similar models for suicide risk outside the USA have also been proposed.[41,42] Comprehensive reviews of these methods have been published previously.[43]

The conditional autoregressive model form used in this study is an ANOVA-style decomposition of the variation in disease risk into separate sets of spatial random effects, temporal random effects and independent space–time interactions (referred to as conditional autoregressive ANOVA hereafter). The ANOVA model form is often used when the aim is to estimate both overall time trends (eg, change in heterogeneity) and spatial patterns (ie, to identify regions of elevated risk).

Additionally, considering intra-state heterogeneity in outcome, and in risk factors and protective factors, we hypothesised that it would be important to build county-level models rather than state-level models.

County-level spatial structure was defined by a binary adjacency matrix, where the adjacency of two counties was inferred from a shared boundary. An analogous temporal adjacency matrix defined the temporal structure. Further details on the methods used and parameter initialisation are in the appendix (pp 6–7).

Statistical analysis

To ascertain that methods that explicitly accommodate spatiotemporal autocorrelation are necessary, we initially built Poisson log-linear models and verified the presence of autocorrelation in their residuals using the Moran's I statistic for spatial and areal data.[44] In the interest of model parsimony, to identify variables with marginal contribution to model quality, we built log-linear models with all possible combinations of predictors considered (2^7-1) and compared their goodness-of-fit (using the Akaike Information Criterion) against that of a model built using all available predictors (appendix pp 9–10). A model with a subset of five predictors (excluding unemployment rate and poverty prevalence variables) was found to have an Akaike Information Criterion close to that of the full model (0·07% larger).

Subsequently, we built conditional autoregressive-ANOVA models with: (1) the full set of predictors; (2) a select subset of predictors; and (3) no covariates (ie, a null model) to measure the predictive skill obtained from using autocorrelation alone. Additionally, we used the expected deaths in a county estimated from differential risk by age, race, and sex of the county's population as a reference model (appendix p 8). This reference model did not capture spatial patterns or temporal trends in suicide mortality and provided a benchmark estimate to assess improvements from conditional autoregressive models. The select model was used as our primary model and all reported findings, unless otherwise stated, are based on estimates derived from that model. Reported effect estimates of the predictors are standardised (ie, per standard deviation). Model convergence was verified using the Geweke and Gelman-Rubin diagnostic tests (appendix p 18).

To estimate annual suicide risk during the study period, we computed mean risk across all counties for each Markov chain Monte Carlo sample, and report median and 95% interval ranges over all samples. Estimates of annual county-specific risk are calculated as the median of all sample estimates for the specific year and county.

Outcome and predictor data overlapped between 2005 and 2016. We defined this time period as our study period, and all models were trained on data for this time period. Since mortality outcome data and data on most of the predictors were available for three additional years (2017–19), these data were used to calculate the out-of-sample predictive model skill, by assuming unavailable predictors remained unchanged since their last known values. For temporal out-of-sample validation, risk for a specific year was predicted with models fit using data up to, but not including, the specific year. For spatial out-of-sample validation, 10-fold cross validation was used: counties were partitioned into ten roughly equal folds and the models were trained in ten iterations. In each iteration, counties in one of the partitions were withheld, models trained with the remaining counties, and the trained models were used to predict risk in withheld counties. The model errors were compared over the fit period (2005–16) and in the two out-of-sample settings:temporal (2017–2019), and spatial (2005–19).

Errors were calculated using the symmetric proportional error (SPE):

$$SPE_{ct} = \frac{Y_{ct} - \hat{Y}_{ct}}{Y_{ct} + \hat{Y}_{ct}},$$

where y_{ct} denotes observed deaths and $\hat{Y}_{ct}$ denotes predicted deaths. SPE has a well defined range and indicates the direction of the error. A division-by-zero issue was avoided by imposing a small lower bound on $\hat{Y}$. An aggregate measure of SPE, the mean SPE, was used to compare model accuracy on both in-sample and out-of-sample predictions. The mean in-sample error for a model was calculated as:

$$\frac{1}{|T_1|C} \sum_{c=1}^{C} \sum_{t \in T_1} |SPE_{ct}|, T_1 = \{2005, ..., 2016\},$$

the mean temporal out-of-sample error as:

$$\frac{1}{|T_2|C} \sum_{c=1}^{C} \sum_{t \in T_2} |SPE_{ct}|, T_2 = \{2017, ..., 2019\},$$

and spatial out-of-sample error as:

$$\frac{1}{|T_1 \cup T_2|C} \sum_{h=1}^{10} \sum_{c \in C_h} \sum_{t \in T_1 \cup T_2} |SPE_{ct}|,$$

where C_h denotes counties in fold h. The Wilcoxon signed-rank test was used to assess statistical significance in the difference of errors for each pair of model forms.[45]

Role of the funding source

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Figure 1 shows the median and 95% credible interval of the posterior effect estimates for each standard deviation change in predictor value (appendix p 16). Median suicide mortality risk increased by 2·8% (95% CrI 1·8 to 3·9) for each 11% increase in state-level firearm ownership rate and by 1·0% (0·4 to 1·5) for each 0·7% increase in state-level prevalence of major depressive disorder. Conversely, median suicide risk decreased by 4·3% (−5·1 to −3·2) for each $12 000 increase in annual median household income in the county, and by 4·3% (−5·8 to −2·5) for each additional 5·8 people per square mile in the county (appendix p 17). Suicide risk decreased by 0·3% (−1·3 to 0·7) for each $162 increase in weekly wage, but this difference was not statistically significant (figure 1A, table; appendix p 16). In the full model, median suicide mortality risk was estimated to decrease by 3·4% (−4·6 to −2·1%) for each 6·4% increase in the prevalence of poverty (figure 1A; appendix p 16) and increased by 2·8% (1·9·3·8%) for each 3% increase in unemployment rate.

Overall, the effect estimates of the variables in the select model remained largely unchanged when additional covariates were introduced (ie, the full model; figure 1A). Both the select and full models indicated strong spatial dependence (ρ_s=0·97) and temporal dependence (ρ_T=0·91; appendix p 17).

The estimated national annual median suicide risk increased from 1·00 (95% CI 0·99–1·02) in 2005 to 1·27 (1·26–1·29) in 2016 (figure 1B).

An increase in the heterogeneity in county-specific suicide risk was also observed during the study period (figure 1C), indicative of a widening gap between counties in which suicide risk is low and counties in which suicide risk is high. Neither trend estimate was found to be sensitive to the set of covariates included in the model (appendix p 20).

The predictive ability of the select model and the full model was comparable (figure 2). The in-sample errors were smaller in the select model than the full model, spatial out-of-sample errors were larger in the select model than the full model, and no significant differences were identified in temporal out-of-sample errors between the two models (p=0·46). This finding implies that two of the covariates did not contribute to model quality in the presence of the other predictors, possibly due to collinearities, yielding a more parsimonious model dependent on fewer data sources.

Error was lower in the select and full models than the null model for all three settings (fit, temporal, and spatial out-of-sample) and the differences were statistically significant (all p<0·05), indicating that the inclusion of covariates improved the model in comparison to accounting for spatiotemporal autocorrelation in suicide mortality alone. Furthermore, all three conditional autoregressive-ANOVA models had smaller errors than the reference model, indicating an improved predictive ability from including spatiotemporal associations. Errors from the reference model were 7–14% larger than errors from the null model (appendix p 21), suggesting that the conditional autoregressive-ANOVA models could be of value even in settings where socioeconomic status predictors of suicide mortality are not available.

Geweke and Gelman-Rubin diagnostic tests (appendix p 18) and visual inspection of trace plots (appendix p 22) indicated model convergence. Analysis of a scatter plot of predicted suicide deaths against observed deaths in the temporal out-of-sample period showed possible overprediction at lower counts (appendix p 23), but otherwise reasonable estimates.

To test whether the conditional autoregressive model was adequate to capture autocorrelation, we ascertained whether autocorrelation was present in model residuals. A visual inspection of the spatial distribution of in-sample residuals (figure 3) and temporal out-of-sample residuals (figure 4) showed no clear spatial structure. No significant spatial autocorrelation in the spatial out-of-sample setting was identified for any year; however, for a majority of the years in both in-sample and temporal out-of-sample settings significant autocorrelation was identified (all p<0·05). The magnitude of the autocorrelation with the conditional autoregressive-ANOVA models was considerably lower than the reference model (appendix p 24).

Discussion

Prediction models of suicide typically only consider clinical characteristics among individuals at high risk of death by suicide in the clinical population and general population; however, suicide risk is also spatially and temporally determined. Our results demonstrate that predictions of death by suicide are improved when aspects of the social environment are used to model risk, and thus aid suicide prevention efforts. Absence of evidence of over-fitting in out-of-sample validation lends confidence to future risk estimates with these models, although more rigorous validation in real-time operational settings is needed.

Considering that the primary objective of this study was to evaluate the feasibility of a predictive model, the covariates considered are not exhaustive. A more comprehensive review of domain literature could help to identify a more robust set of predictors, including aspects of the physical environment that are not collinear with the predictors considered.[42] Additional stratifications of the predictors used or other socioeconomic status indicators, such as measures of social cohesion, access to health care, unemployment rates in specific sectors of the economy, and prevalence of depression stratified by age, prescription rates for antidepressant and pain management medications, and housing quality (crowded living conditions, access to green spaces, ambient vehicular noise) could also be tested. When used at the sub-national scale (ie, state level or local public health agency level), additional predictors of local relevance that do not have national coverage might also be viable. Previous studies have reported strength of associations of different suicide risk factors throughout the life course and appropriate interventions for different population subgroups.[46,47]

The trend in national mean risk is consistent with increases in mortality rates reported by multiple studies for the USA overall and in almost all demographic groups.[1]

County-specific risk estimates can help categorise counties—eg, counties with relatively stable risk especially those that remained in the highest or lowest deciles, or counties in which the largest year-to-year changes were observed—and hence help identify areas in greater need of preventive resources, or conversely identify areas where interventions seem to be effective.

The protective effect of poverty in the full model (3·4% decrease in suicide risk per each 6·4% increase in county-level prevalence of poverty) needs more careful examination and interpretation, ideally in conjunction with measures of rurality, societal fragmentation, and poverty persistence[22] and interactions with other predictors in the model. The findings are not anomalous since previous studies have reported mixed associations between poverty prevalence and suicide rates at the population level, with results varying by study design, including geographical resolution and population strata analysed.[20] At the individual level, the negative impact of poverty on suicide rates is more consistent across studies.[48]

The predictions from the models were at annual resolution and hence not responsive to real-time changes in risk. Although the model structure does not preclude generation of weekly or monthly risk estimates, a barrier to such an operational deployment is the paucity of reliable, timely measures of suicidal activity (thoughts or attempts). Identification of

sources for near real-time situational data, and development of nowcast models to translate data feeds into measures of suicidal ideation in the community, could substantially aid the translation of the models into a surveillance setting.

The volume of calls to crisis hotline centres could represent a potential source of reliable information on suicidal ideation. A national suicide prevention and mental health crisis number (9-8-8) was lauched in the USA on July 16, 2022, representing a unified system with wide coverage. Similarly, with the deployment of the National Emergency Medical Services Information System, timely information on Emergency Medical Service requests is also available.[49] Aggregate event data at the county level from these or similar sources, if made publicly available every week, could support the development of nowcast systems. Such models are in use for numerical weather prediction,[50] macroeconomic analyses,[51] and influenza surveillance,[52,53] among other domains.

Posts to thematically related social media sites,[54,55] queries on search engines[56] or access logs to suicide prevention forums and related websites might also serve as valuable indicators of suicidal crisis or ideation. Measures of ideation among adolescents and young adults, among whom suicide is in the top three most common causes of death, could be particularly valuable.[57] Initial results from nowcast models of suicide mortality using crisis call volumes and online search activity have been encouraging (data not shown).

Limitations of the study include ad-hoc independent variable selection, the assumption of geographical homogeneity for predictors for which county-level data were not available, a potentially simplistic spatial adjacency matrix that did not reflect population mobility and mixing nor online social interconnectedness, inflated error estimates in counties with no deaths by suicide in a year (appendix p 11), and potential underestimates of suicide deaths and risk in certain racial or ethnic groups due to inconsistent suicide certification practices.[58]

The reduction, but not elimination, of spatial autocorrelation could be indicative of the insufficiency of the specific conditional autoregressive model form used or of the covariates considered. Additionally, two of the predictors—prevalence of major depressive episode and firearm ownership—were only available at state-wise resolution, a shortcoming addressed by assuming all counties in a state to have identical values, which may have contributed to an increase in residual spatial autocorrelation.

Reducing suicide deaths in addition to unintentional drug overdose deaths, is crucial to public health efforts aimed at reversing declines in life expectancy in the USA. Coupled with measures of situational awareness, conditional autoregressive models have the capacity to flag anomalous changes in suicide risk at the population level. This could inform the timely deployment of targeted interventions in locations most at need, and prompt investigations if increases are concentrated in specific population groups. Operationalisation of such surveillance systems by public health agencies would require access to timely county-level data, of both suicide mortality and risk or protective factors, potentially supplemented by reliable proxies of suicidal behaviour, in addition to support for personnel to implement and maintain model codebase and related data infrastructure. These efforts can be accelerated

with improved infrastructure for more rapid and granular data surveillance on factors such as depression and mental health in the population, and access to lethal means.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data sharing

Covariates were gathered from public datasets and references to the primary sources have been provided. Mortality data was acquired under a restricted use agreement that does not allow sharing data publicly. R scripts to run the models and perform the analysis, and sample data and results can be accessed online.

References

1. Hedegaard H, Curtin SC, Warner M. Increase in suicide mortality in the United States, 1999–2018. NCHS Data Brief 2020; 362: 1–8.

2. Woolf SH, Schoomaker H. Life expectancy and mortality rates in the United States, 1959–2017. JAMA 2019; 322: 1996–2016. [PubMed: 31769830]

3. Gradus JL, Rosellini AJ, Horváth-Puhó E, et al. Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. JAMA Psychiatry 2020; 77: 25–34. [PubMed: 31642880]

4. Roaten K, Johnson C, Genzel R, Khan F, North CS. Development and implementation of a universal suicide risk screening program in a safety-net hospital system. Jt Comm J Qual Patient Saf 2018; 44: 4–11. [PubMed: 29290245]

5. Belsher BE, Smolenski DJ, Pruitt LD, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA Psychiatry 2019; 76: 642–51. [PubMed: 30865249]

6. Gradus JL, Rosellini AJ, Horváth-Puhó E, et al. Predicting sex-specific non-fatal suicide attempt risk using machine learning and data from Danish national registries. Am J Epidemiol 2021; 190: 2517–27 [PubMed: 33877265]

7. Gunnell D, Knipe D, Chang S-S, et al. Prevention of suicide with regulations aimed at restricting access to highly hazardous pesticides: a systematic review of the international evidence. Lancet Glob Health 2017; 5: e1026–37 [PubMed: 28807587]

8. Miller M, Lippmann SJ, Azrael D, Hemenway D. Household firearm ownership and rates of suicide across the 50 United States. J Trauma 2007; 62: 1029–34, discussion 1034–35. [PubMed: 17426563]

9. Siegel M, Rothman EF. Firearm ownership and suicide rates among US men and women, 1981–2013. Am J Public Health 2016; 106: 1316–22. [PubMed: 27196643]

10. Baker ST, Nicholas J, Shand F, Green R, Christensen HJAP. A comparison of multi-component systems approaches to suicide prevention. Australas Psychiatry 2018; 26: 128–31. [PubMed: 29160089]

11. Blanco C, Wall MM, Olfson M. A population-level approach to suicide prevention. JAMA 2021; 325: 2339–40. [PubMed: 34042948]

12. Blakely TA, Collings SC, Atkinson J. Unemployment and suicide. Evidence for a causal association? J Epidemiol Community Health 2003; 57: 594–600. [PubMed: 12883065]

13. Lewis G, Sloggett A. Suicide, deprivation, and unemployment: record linkage study. BMJ 1998; 317: 1283–86. [PubMed: 9804714]

14. Kegler SR, Stone DM, Holland KM. Trends in suicide by level of urbanization—United States, 1999–2015. MMWR Morb Mortal Wkly Rep 2017; 66: 270–73. [PubMed: 28301448]

15. Canetto SS, Sakinofsky I. The gender paradox in suicide. Suicide Life Threat Behav 1998; 28: 1–23. [PubMed: 9560163]

16. Conwell Y, Van Orden K, Caine ED. Suicide in older adults. Psychiatr Clin North Am 2011; 34: 451–68. [PubMed: 21536168]

17. Rihmer Z. Suicide risk in mood disorders. Curr Opin Psychiatry 2007; 20: 17–22. [PubMed: 17143077]

18. Sokero TP, Melartin TK, Rytsälä HJ, Leskelä US, Lestela-Mielonen PS, Isometsä ET. Prospective study of risk factors for attempted suicide among patients with DSM-IV major depressive disorder. Br J Psychiatry 2005; 186: 314–18. [PubMed: 15802688]

19. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017; 143: 187–232. [PubMed: 27841450]

20. Rehkopf DH, Buka SL. The association between suicide and the socio-economic characteristics of geographical areas: a systematic review. Psychol Med 2006; 36: 145–57. [PubMed: 16420711]

21. The National Research Council. Firearms and violence: a critical review. Washington, DC: National Academies Press, 2005.

22. Congdon P. The spatial pattern of suicide in the US in relation to deprivation, fragmentation and rurality. Urban Stud 2011; 48: 2101–22. [PubMed: 22069804]

23. Durkheim E. Suicide: a study in sociology. Glencoe, IL: The Free Press, 1951.

24. Phillips JA. Factors associated with temporal and spatial patterns in suicide rates across U.S. states, 1976–2000. Demography 2013; 50: 591–614. [PubMed: 23196429]

25. Cheng Q, Li H, Silenzio V, Caine ED. Suicide contagion: a systematic review of definitions and research utility. PLoS One 2014; 9: e108724. [PubMed: 25259604]

26. Koda M, Harada N, Eguchi A, Nomura S, Ishida Y. Reasons for suicide during the COVID-19 pandemic in Japan. JAMA Network Open 2022; 5: e2145870. [PubMed: 35099547]

27. Niederkrotenthaler T, Braun M, Pirkis J, et al. Association between suicide reporting in the media and suicide: systematic review and meta-analysis. BMJ 2020; 368: m575. [PubMed: 32188637]

28. Sisask M, Värnik A. Media roles in suicide prevention: a systematic review. Int J Environ Res Public Health 2012; 9: 123–38. [PubMed: 22470283]

29. Keyes KM, Kandula S, Olfson M, et al. Suicide and the agent-host-environment triad: leveraging surveillance sources to inform prevention. Psychol Med 2021; 51: 529–37 [PubMed: 33663629]

30. Sy KTL, Shaman J, Kandula S, Pei S, Gould M, Keyes KM. Spatiotemporal clustering of suicides in the US from 1999 to 2016: a spatial epidemiological approach. Soc Psychiatry Psychiatr Epidemiol 2019; 54: 1471–82. [PubMed: 31177308]

31. Centers for Disease Control and Prevention, National Center for Health Statistics. National Vital Statistics System. Restricted-use Vital Statistics Data. https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm (accessed Jan 5, 2023).

32. Centers for Disease Control and Prevention, National Center for Health Statistics. National Vital Statistics System. Bridged-race categories. https://www.cdc.gov/nchs/nvss/bridged_race.htm (accessed Jan 5, 2023).

33. Centers for Disease Control and Prevention, National Center for Health Statistics. Vintage 2020 bridged-race postcensal population estimates. https://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm (accessed Feb 17, 2022).

34. United States Census Bureau. SAIPE State and County Estimates for 2020. https://www.census.gov/data/datasets/2020/demo/saipe/2020-state-and-county.html (accessed Jan 5, 2023).

35. Substance Abuse and Mental Health Data Archive. National Survey on Drug Use and Health (NSDUHs). Population data. https://www.datafiles.samhsa.gov/dataset/national-survey-drug-use-and-health-2019-nsduh-2019-ds0001 (accessed Jan 5, 2023).

36. Schell TL, Peterson S, Vegetabile BG, Scherling A, Smart R, Morral AR. State-level estimates of household firearm ownership. Santa Monica, CA: RAND, 2020. https://www.rand.org/pubs/tools/TL354.html (accessed Jan 5, 2023).

37. US Bureau of Labor Statistics. Quarterly census of employment and wages. https://www.bls.gov/cew/downloadable-data-files.htm (accessed Jan 5, 2023).

38. US Bureau of Labor Statistics. Local area unemployment statistics. 2021. https://www.bls.gov/lau/lauov.htm (accessed Nov 18, 2022).

39. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 1991; 43: 1–20.

40. Leroux BG, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran EM, Berry D. Statistical models in epidemiology, the environment, and clinical trials. New York, NY: Springer, 2000: 179–91.

41. Helbich M, Plener PL, Hartung S, Blüml V. Spatiotemporal suicide risk in Germany: a longitudinal study 2007–11. Sci Rep 2017; 7: 7673. [PubMed: 28794489]

42. Helbich M, de Beurs D, Kwan M-P, O'Connor RC, Groenewegen PP. Natural environments and suicide mortality in the Netherlands: a cross-sectional, ecological study. Lancet Planet Health 2018; 2: e134–39. [PubMed: 29546252]

43. Lawson AB. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. London: Chapman and Hall, 2018.

44. Moran PA. Notes on continuous stochastic phenomena. Biometrika 1950; 37: 17–23. [PubMed: 15420245]

45. Hollander M, Wolfe DA. Nonparametric statistical methods, 3rd edn. Chichester: Wiley, 1999.

46. Fazel S, Runeson B. Suicide. N Engl J Med 2020; 382: 266–74. [PubMed: 31940700]

47. Knipe D, Padmanathan P, Newton-Howes G, Chan LF, Kapur N. Suicide and self-harm. Lancet 2022; 399: 1903–16. [PubMed: 35512727]

48. Olfson M, Cosgrove C, Altekruse SF, Wall MM, Blanco C. Deaths of despair: adults at high risk for death by suicide, poisoning, or chronic liver disease in the US: study examines US adults at highest risk for death by suicide, drug poisoning, or chronic liver disease. Health Aff 2021; 40: 505–12.

49. Hallowell BD, Chambers LC, Rhodes J, Basta M, Viner-Brown S, Lasher L. Using emergency medical services data to monitor nonfatal opioid overdoses in real time: development, validation, and use of a case definition, Rhode Island, 2018. Public Health Rep 2021; 136 (suppl): 40S–46S. [PubMed: 34726979]

50. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-C. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Adv Neural Inf Process Syst 2015; 1: 802–10.

51. Ba bura M, Giannone D, Modugno M, Reichlin L. Now-casting and the real-time data flow. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1564.pdf (accessed Jan 5, 2023).

52. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009; 457: 1012–14. [PubMed: 19020500]

53. Kandula S, Shaman J. Reappraising the utility of Google flu trends. PLoS Comput Biol 2019; 15: e1007258. [PubMed: 31374088]

54. Eichstaedt JC, Smith RJ, Merchant RM, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci USA 2018; 115: 11203–08. [PubMed: 30322910]

55. O'dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015; 2: 183–88.

56. Lee J-Y. Search trends preceding increases in suicide: a cross-correlation study of monthly Google search volume and suicide rate using transfer function models. J Affect Disord 2020; 262: 155–64. [PubMed: 31733460]

57. Heron M. Deaths: leading causes for 2019. https://www.cdc.gov/nchs/data/nvsr/nvsr70/nvsr70-09-508.pdf (accessed Jan 5, 2023)..

58. Rockett IR, Wang S, Stack S, et al. Race/ethnicity and potential suicide misclassification: window on a minority suicide paradox? BMC Psychiatry 2010; 10: 35 [PubMed: 20482844]

Research in context

Evidence before this study

To identify relevant studies, we searched PubMed from database inception to May 10, 2022 using the terms "US", "county", "suicide mortality", and "socioeconomic'", without language restrictions. The search identified five articles that were reviewed to assess relevance and summarise findings. Two of the studies were deemed not directly relevant since they were limited to a single county or specific to suicides following a cancer diagnosis. None of the studies explicitly accounted for spatiotemporal autocorrelation. Widening the search terms to "US'", "suicide mortality", and "risk factors" identified around 50 association studies on individual risk factors for suicide, including characteristics such as age, race or ethnicity, sex, socioeconomic status, and mental health history. Studies assessing the combined effect sizes of these characteristics were less common and meta-analyses of effect sizes demonstrated considerable heterogeneity. Spatiotemporal models for suicide risk outside the USA have been previously reported.

Added value of this study

Although understanding the direction and magnitude of effects is essential, our focus was on predicting future suicide risk. The proposed model estimates county-level suicide risk while explicitly accounting for spatiotemporal autocorrelation, yielding geographically well resolved, and arguably more actionable, estimates.

Implications of all the available evidence

Building on previous findings on risk factors for suicide, our results demonstrate that predictions of suicide deaths at population scale are improved when aspects of the social environment are used to model risk. Operational systems to flag changes in suicide risk in a community would benefit from timely data on suicide mortality, mental health, and related risk factors.
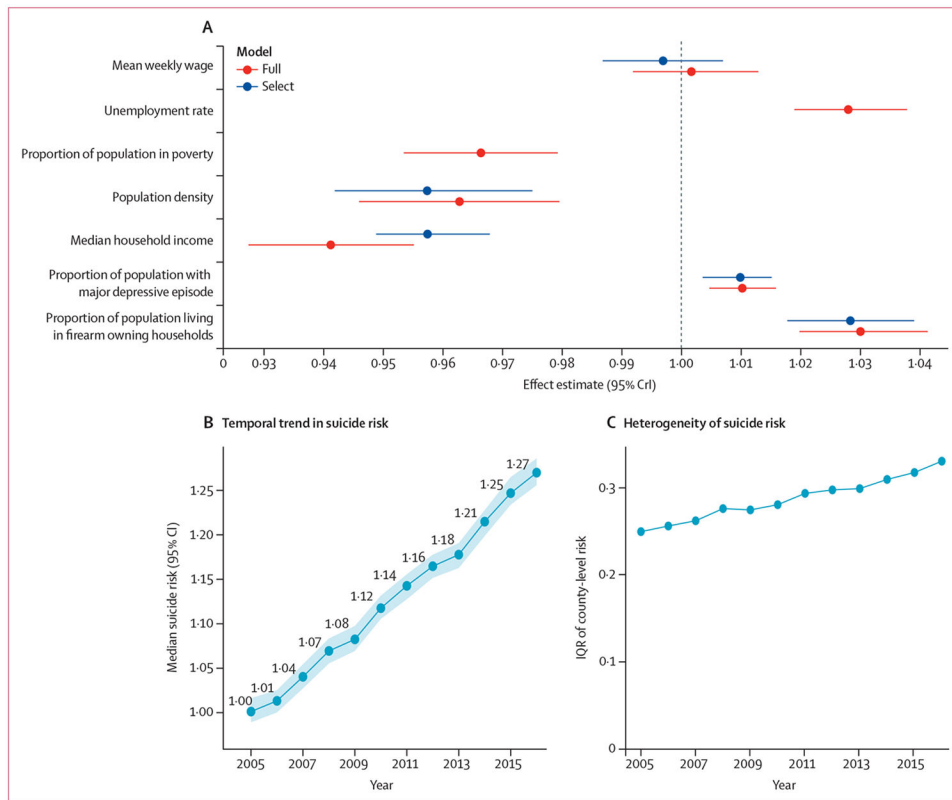
Figure 1: Effect estimates and trends in suicide risk (2005–16)

(A) Posterior median and 95% CrIs of the effect estimates from conditional autoregressive-ANOVA models, per one SD change in variable. (B) Estimates of national suicide risk; shaded area shows 95% CIs. (C) IQR of county-level risk as estimated by the select model; higher IQRs indicate greater heterogeneity and an increase in IQR with time indicates a widening gap between counties in which suicide risk is low and counties in which suicide risk is high. CrI=credible interval.
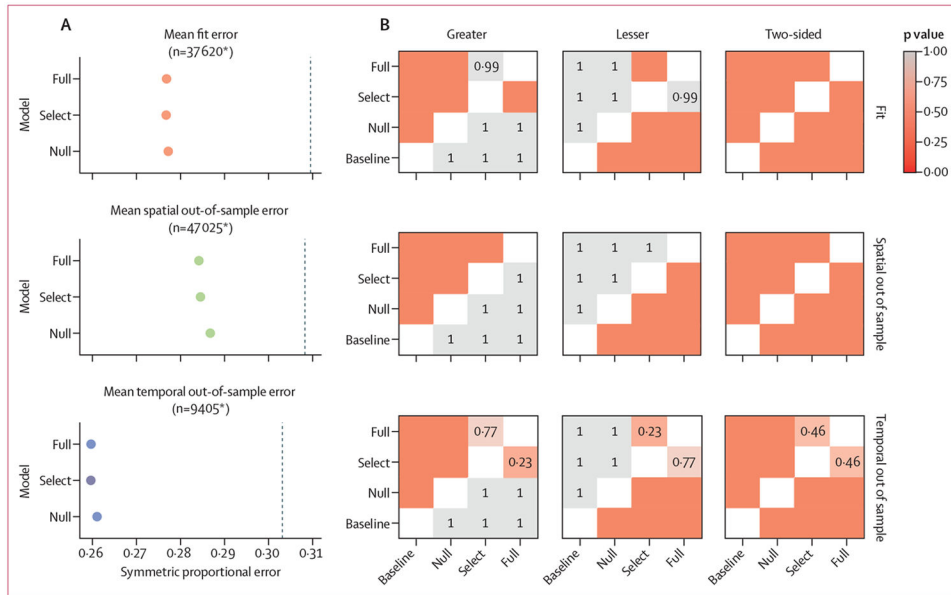
Figure 2: Comparison of error by model and setting

(A) Mean symmetric proportional error for in-sample, temporal out-of-sample, and spatial out-of-sample estimates. The vertical dashed line denotes error from the reference model. (B) p values derived from the Wilcoxon signed-rank test for each pair of models. Significance ($p<0.05$) in the two-sided column indicates that the difference in errors of model X (x-axis) and model Y (y-axis) is not symmetric around 0. p values of <0.05 in the lesser panel column indicates errors in model X (x-axis) are lower than model Y (y-axis). Actual p values are shown for errors that were not significantly different. To check for statistically significant differences in errors between any pair of models, we checked for significance with a two-sided test and if significant, checked for significance under one of lesser or greater panel columns. *Number of county-year combinations of suicide risk across which mean is reported.

Figure 3: Symmetric proportional in-sample errors for the select model (2005–16)
Negative error indicates model estimates exceeded observed suicide deaths. Counties with
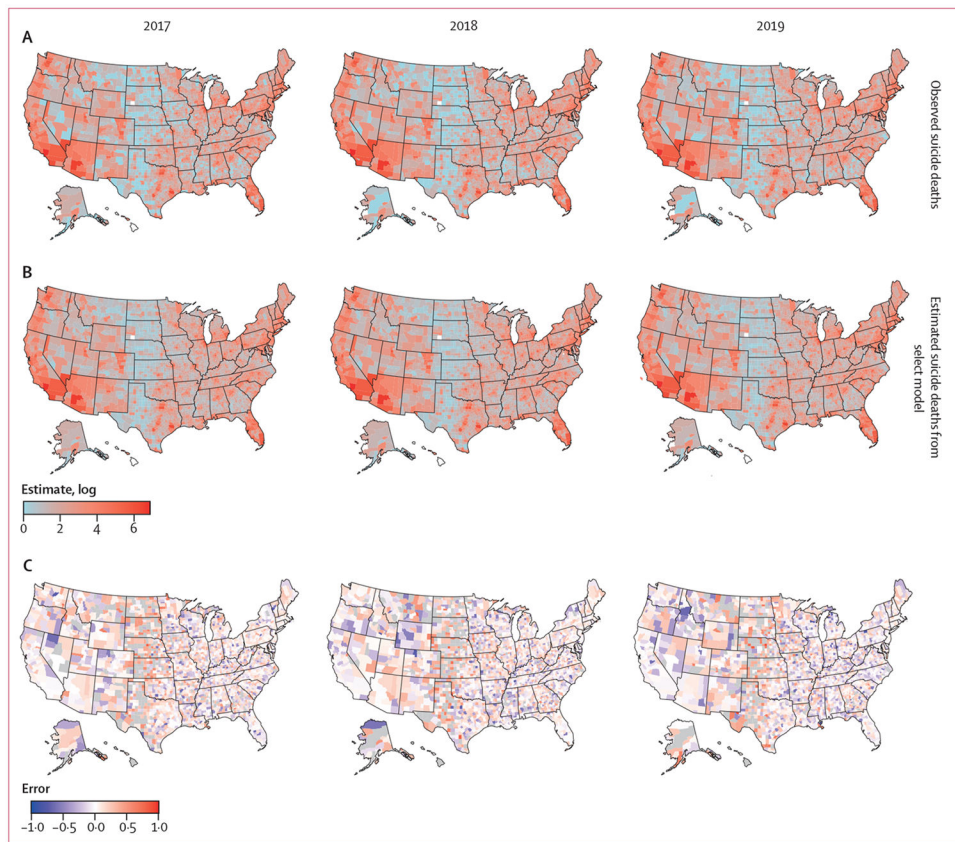no observed deaths are shown in grey.

Figure 4: Observed suicide deaths (A), out-of-sample model estimates of suicide deaths (B), and symmetric proportional errors (C) for the select model (2017–19)

Negative error indicates model estimate exceeds observed. Counties with no observed deaths are shown in grey.

Table:

Explanatory variables used in the study

	Data source	Available period	Geographical resolution	Mean (SD)
Observed suicide-related deaths	NVSS	2003–19	County level	12·3 (32)
Unemployment rate	BLS	2003–19	County level	0·067 (0·03)
Mean weekly wage, $US	BLS	2003–19	County level	664 (162)
Proportion of population living in poverty	SAIPE	2003–19	County level	0·162 (0·06)
Median household income, US$	SAIPE	2003–19	County level	44 392 (11 749)
Proportion of population living in firearm owning households	RAND	2005–16	State level	0·412 (0·11)
Proportion of population with major depressive episode	NSDUH	2005–18	State level	0·069 (0·01)
Population density per square mile, log	Census	2005–19	County level	3·77 (1·8)

Mean and SDs are reported for 2005–16 for all variables. NVSS=National Vital Statistics System. BLS=Bureau of Labor Statistics. SAIPE=Small Area Income and Poverty Estimates. RAND=The RAND Corporation. NSDUH=National Surveys on Drug Use and Health.