WEB MATERIAL

Geographic Variation, Economic Activity, and Labor Market Characteristics in Trajectories of Suicide in

the United States, 2008–2020

Katherine M. Keyes, Sasikiran Kandula, Gonzalo Martinez-Ales, Catherine Gimbrone, Victoria Joseph, Shannon Monnat, Caroline Rutherford, Mark Olfson, Madelyn Gould, and Jeffrey Shaman

Contents:

Web Figure 1. Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and log likelihood of GMM models of 3,140 counties at different number of clusters (x-axis) for different responses (columns). The vertical dashed line denotes the selected model, *k=5*. Base model (red) and model with covariates (blue) have very similar goodness-of-fit.
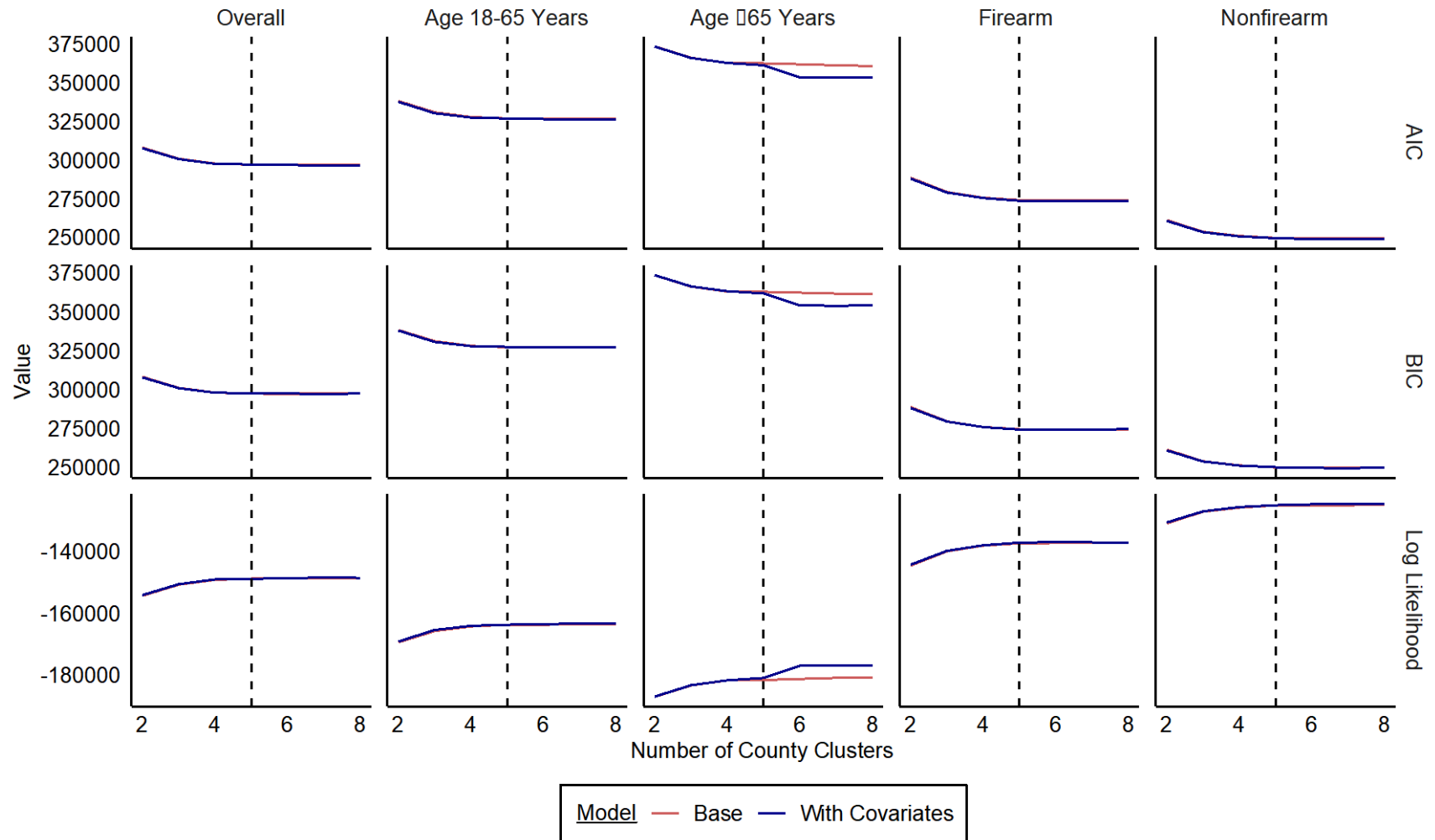
Web Figure 2. Relative entropy of GMM base models (red) and model with covariates (blue) at different number of clusters (x-axis) for different responses (subpanels). There were 3,140 counties included in the models. The horizontal dashed line denotes the acceptable threshold of 0.5.
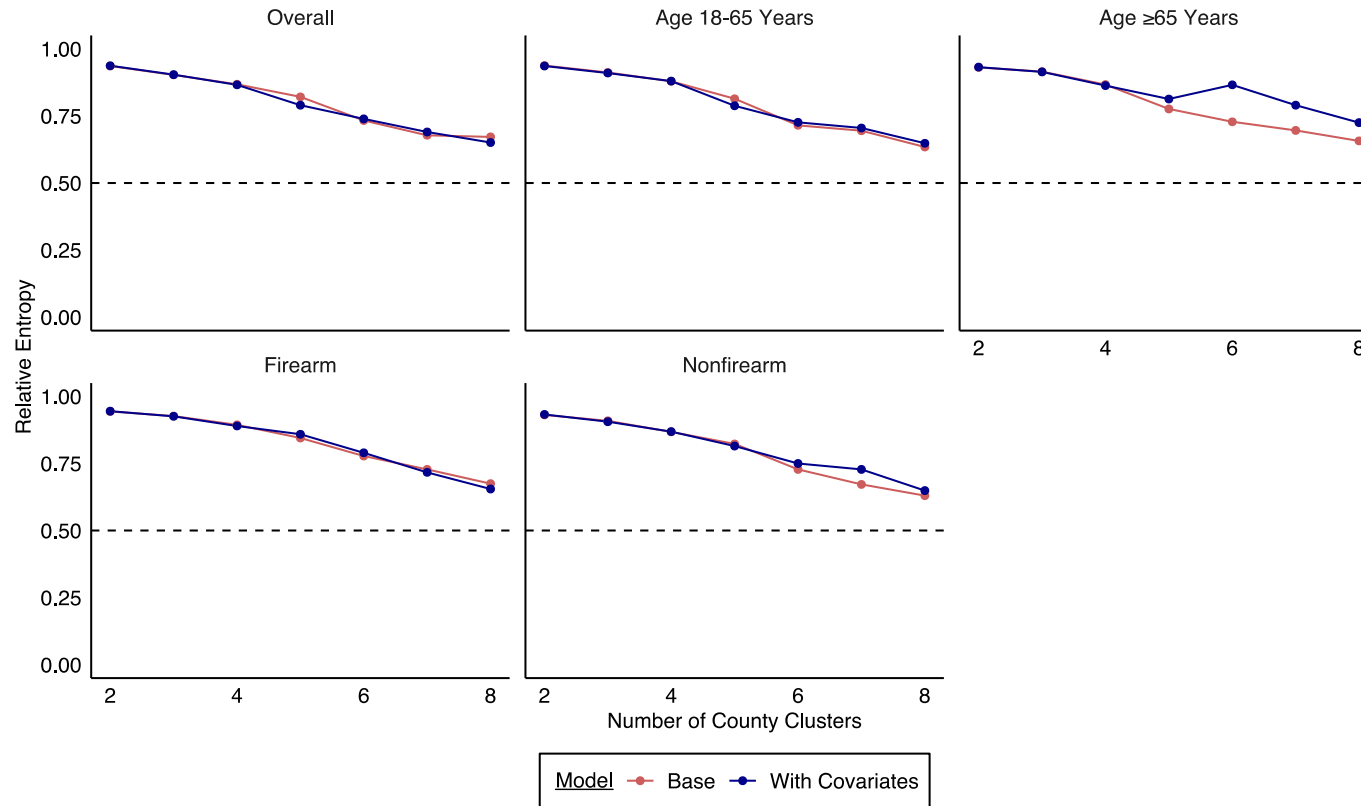

The entropy of a model with K latent classes is calculated as:

$$E(K) = -\sum_{k=1}^{K}\sum_{i=1}^{N} \hat{p}_{ik} \ln(\hat{p}_{ik})$$

where $\hat{p}_{ik}$, is as defined above and N is the total number of counties. Entropy, unlike APPA, is a global measure defined across all classes, with values close to 0 indicating low classification uncertainty, and an undefined upper bound (i.e. bounds of $[0, \infty)$). Relative entropy rescales the values to $[0, 1]$ and is calculated as $1 - \frac{E(K)}{n \log(K)}$, i.e. higher values of relative entropy indicate higher classification certainty and values greater than 0.5 are considered acceptable (43,52,54,55).

Web Figure 3. Class-specific Average Posterior Probability of Assignment (APPA) of GMM base models (red) and model with covariates (blue) at different number of clusters (columns) for different responses (rows). There were 3,140 counties included in the models. The horizontal dashed line denotes the acceptable threshold of 0.7.
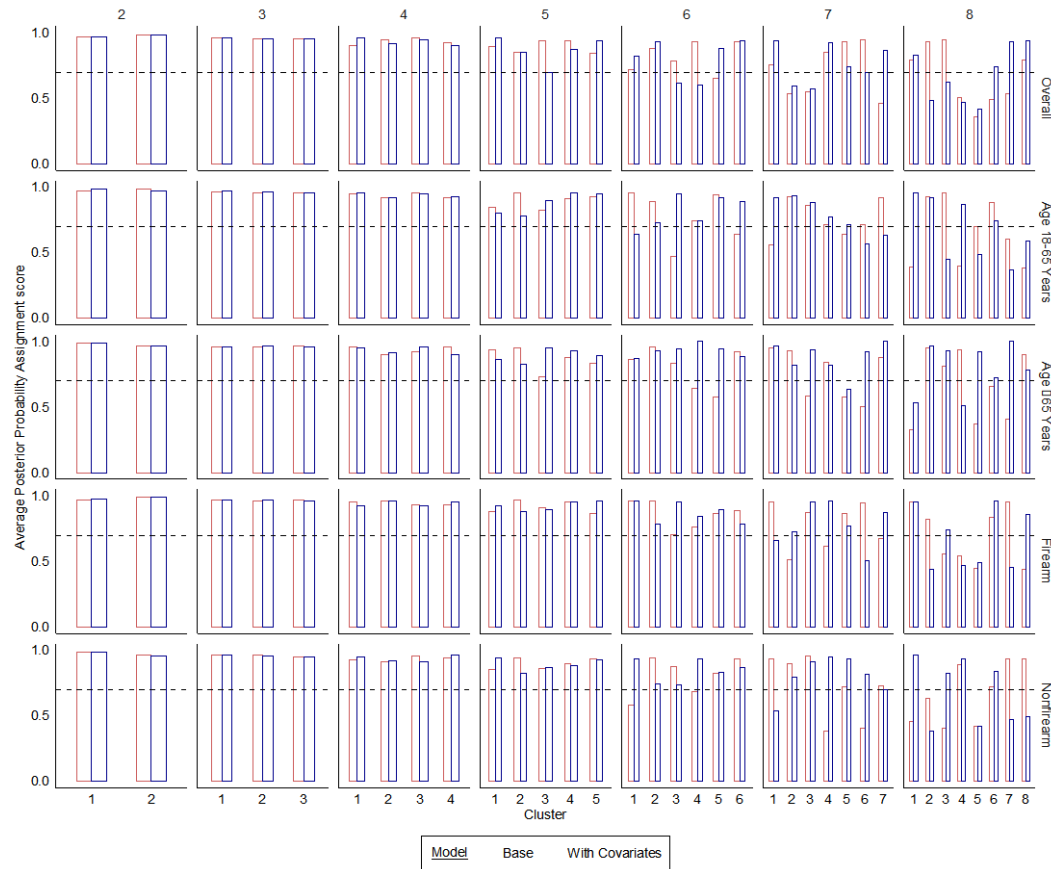

Let $\hat{p}_{ik}$ be the posterior probability of county i to belong to class k, as determined by the model; hence $\sum_{k=1}^{K} \hat{p}_{ik} = 1$. A county is assumed to be *assigned* to the class with the maximum $\hat{p}_{ik}$. If n_k is the number of counties assigned to class k, the APPA of cluster k is calculated by summing the posterior probabilities of all counties assigned to class k:

$$APPA_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \hat{p}_{ik}$$

In other words, APPA calculates the average posterior probability of all counties assigned to class k. APPA has clear bounds, [0, 1], with values closer to 1 indicating greater classification certainty; an APPA greater than 0.7 for *all* classes is considered acceptable (43,52,53).

Web Figure 4. (*top*) Mean trajectories of posterior estimates of county suicide rates among 18-65 year population (*left*) and among 65+ population (*right*), per the 5-class GMM *Overall* model of 3,140 counties. Text at the end of the trajectory indicates the percentage of counties assigned to the class. Classes are ordered by the 2008 average rate of suicide death of counties assigned to the cluster (least risk to most risk); (*bottom*) Corresponding maps of county class assignment by the 5-class GMM model.
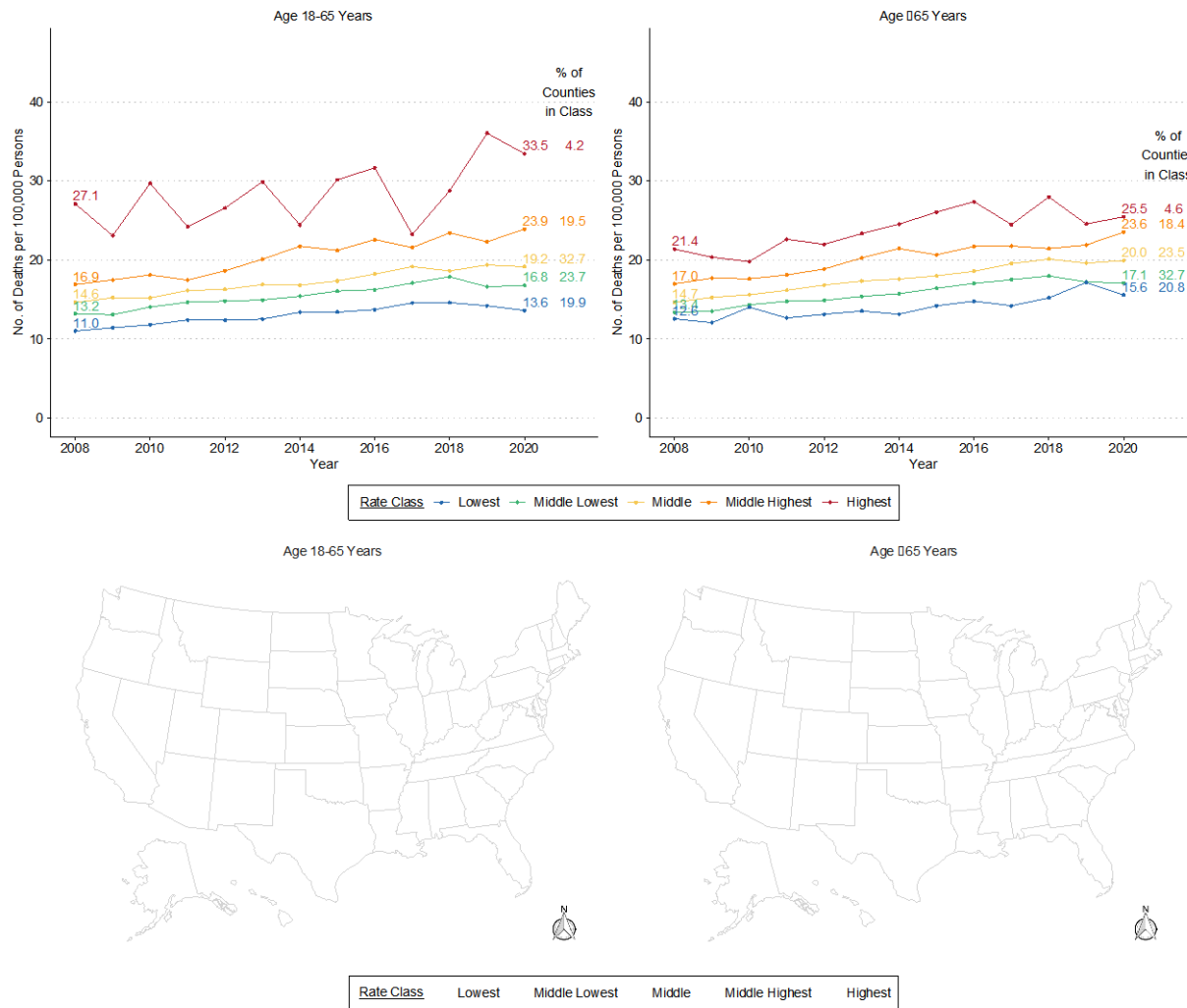
Web Figure 5. (*top*) Mean trajectories of posterior estimates of county suicide rates involving firearms (*left*) and not involving firearms (*right*), per the 5-class GMM *Overall* model of 3,140 counties. Text at the end of the trajectory indicates the percentage of counties assigned to the class. Classes are ordered by the 2008 average rate of suicide death of counties assigned to the cluster (least risk to most risk); (*bottom*) Corresponding maps of county class assignment by the 5-class GMM model.

Web Table 1. Average county demographic characteristics* and characteristics of suicide deaths* by suicide rate cluster in 2008 and 2020

Variable	2008					2020				
	Average Suicide Rate					Average Suicide Rate				
	Lowest	Middle Lowest	Middle	Middle Highest	Highest	Lowest	Middle Lowest	Middle	Middle Highest	Highest
County demographics										
% Male	49.1	49.75	50.19	50.53	50.92	49.17	49.89	50.39	50.84	51.05
% White	81.97	86.53	88.17	92.04	86.52	79.38	85.34	86.99	90.28	84.54
% Black	13.17	11.18	9.04	4.02	2.16	14.47	11.69	9.43	4.59	3.12
% Asian/Pacific Islander	3.83	1.03	0.66	0.53	0.43	4.95	1.47	1.06	0.99	0.87
% Native American / Indigenous	1.02	1.26	2.13	3.41	10.9	1.2	1.51	2.51	4.14	11.47
% Hispanic	11.64	6.46	7.04	8.13	8.2	14.07	8.16	8.87	9.98	10.66
Average population density**	1017.71	121.72	64.56	23.5	2.17	1074.18	127.19	65.87	23.26	2.1
Characteristics of suicide deaths										
% Male	78.13	79.39	69.82	53.42	33.72	79	80.05	75.52	62.16	37.23
% White	90.65	91.69	78.93	60.65	27.94	86.98	90.63	83.45	67.57	34.4
% Black	6.06	4.32	3.15	1.08	0.81	8.4	5.62	3.96	1.04	1.01
% Asian/Pacific Islander	1.8	0.5	0.17	0.1	0	2.85	0.78	0.53	0.07	0
% Native American / Indigenous	0.52	0.72	1.91	2.48	8.35	0.8	1.31	2.13	3.51	8.14
% Hispanic	5.52	3.25	2.85	2.79	2.07	8.12	4.69	4.88	3.6	1.75
% Not Hispanic	93.31	93.72	81.05	61.3	35.02	90.74	93.39	85.07	68.58	41.8
% Hispanic origin not stated	0.2	0.26	0.26	0.23	0	0.17	0.26	0.12	0	0

* Demographics and characteristics of suicide deaths estimated as the cluster average of the average of each county within the cluster. For example, we extracted the percent male for each county in the cluster, and estimated the average of those percentages across counties. Thus, column percentages do not sum to 100 percent.

** People per square mile.