

1 **Risk factors associated with mortality of COVID-19**
2 **in 3125 counties of the United States**

3 Ting Tian^{1#}, Jingwen Zhang^{1#}, Liyuan Hu^{1#}, Yukang Jiang^{1#}, Congyuan Duan¹,
4 Zhongfei Li⁵, Xueqin Wang^{2,3*}, Heping Zhang^{4*}

5 **Affiliations:**

6 ¹School of Mathematics, Sun Yat-sen University, Guangzhou, Guangdong, China

7 135 Xingang Xi Road, Guangzhou, Guangdong, 510275, China

8 ²School of Statistics, Capital University of Economics and Business, Beijing, China

9 121 Zhangjialukou, Huaxiang Fengtai District, Beijing, 100070, China

10 ³School of Management, University of Science and Technology of China, Hefei,

11 Anhui, China

12 96 Jinzhai Road, Hefei, Anhui, 230026, China

13 ⁴School of Public Health, Yale University, New Haven, CT, USA

14 60 College Street, New Haven, CT, 06520-8034, USA

15 ⁵School of Management, Sun Yat-sen University, Guangzhou, Guangdong, China

16 135 Xingang Xi Road, Guangzhou, Guangdong, 510275, China

17

18 [#]Dr. T. Tian, J. Zhang, L. Hu and Y. Jiang contributed equally to this article

19 ^{*}Dr. X. Wang and Dr. H. Zhang are corresponding authors

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

20

21 **Abstract**

22 **Background**

23 The number of cumulative confirmed cases of COVID-19 in the United States has
24 risen sharply since March 2020. A county health ranking and roadmaps program has
25 been established to identify factors associated with disparity in mobility and mortality
26 of COVID-19 in all counties in the United States.

27 **Methods**

28 To find out the risk factors associated with county-level mortality of COVID-19 with
29 various levels of prevalence, a negative binomial design was applied to the county-
30 level mortality counts of COVID-19 as of August 27, 2020 in the United States. In
31 this design, the infected counties were categorized into three levels of infections using
32 clustering analysis based on time-varying cumulative confirmed cases from March 1
33 to August 27, 2020. COVID-19 patients were not analyzed individually but were
34 aggregated at the county-level, where the county-level deaths of COVID-19
35 confirmed by the local health agencies.

36 **Results**

37 3125 infected counties were assigned into three classes corresponding to low, median,
38 and high prevalence levels of infection. Several risk factors were significantly
39 associated with the mortality counts of COVID-19, where higher level of air pollution
40 (0.153, $P < 0.001$) increased the mortality in the low prevalence counties and elder

41 individuals were more vulnerable in both the median and high prevalence counties .

42 The segregation between non-Whites and Whites and higher Hispanic population had

43 higher likelihood of risk of the deaths in all infected counties.

44 **Conclusions**

45 The mortality of COVID-19 depended on sex, race/ethnicity, and outdoor

46 environment. The increasing awareness of the impact of these significant factors may

47 lead to the reduction in the mortality of COVID-19.

48

49 **Key words:** adverse health factors; county-level confirmed and deaths; race/ethnicity;

50 segregation index; physical environment

51

52

53

54

55

56

57

58

59

60

61

62 **Introduction**

63 COVID-19 is an infectious disease caused by a novel coronavirus with an
64 estimated average incubation period of 5.1 days[1]. It spreads through person-to-
65 person transmission, and has now infected 215 countries and regions with over 24
66 million total confirmed cases as of August 27, 2020[2]. The United States had
67 5,867,785 confirmed cases on August 27, 2020, the highest in the world, but there
68 were only 69 confirmed cases on March 1, 2020[3].

69 The United States has been suffering from a severe epidemic, with COVID-19
70 related deaths occurring all over the country. For instance, New York City had the
71 largest number of total deaths (23,674), accounting for the majority of deaths in the
72 infected counties, while no one in King county, Texas was infected as of August 27,
73 2020[3]. Therefore, it is of great interest to find out the risk factors that influence the
74 number of deaths of COVID-19. It is known that infectious diseases are affected by
75 factors other than medical treatments[4, 5]. For example, influenza A is associated
76 with obesity[6], and the spread of the 2003 SARS events depends on seasonal
77 temperature changes[7].

78 The County Health Rankings and Roadmaps program was launched by both the
79 Robert Wood Johnson Foundation and the University of Wisconsin Population Health
80 Institute[8]. This program has been providing annual sustainable source data
81 including health outcomes, health behaviors, clinical care, social and economic
82 factors, physical environment and demographics since 2010. We explored putative

83 risk factors that may affect the mortality of COVID-19 in different areas of the United
84 States in order to increase awareness of the disparity and aid the development of risk
85 reduction strategies.

86

87 **Methods**

88 **Data sources**

89 We collected the number of cumulative confirmed cases and deaths from March
90 1 to August 27, 2020, for counties in the United States from the New York Times[9].
91 The COVID-19 confirmed cases and deaths were identified by the laboratory RNA
92 test and specific criteria for symptoms and exposures from health departments and
93 U.S. Centers for Disease Control and Prevention (CDC). The county health rankings
94 reports from year 2020 were compiled from the County Health Rankings and
95 Roadmaps program official website[8]. There were 77 measures in each of 3142
96 counties, including the health outcome, health behaviors, clinical care, social and
97 economic factors, physical environment, and demographics. We refer to the official
98 website of the County Health Rankings and Roadmaps program[8] for detailed
99 information.

100

101 **Study areas**

102 As of August 27, 2020, a total of 3,208 counties reported confirmed cases in the
103 United States, leaving 3125 counties with both confirmed cases of COVID-19 and

104 county health ranking data recorded to be analyzed in this study. The total number of
105 deaths as of August 27, 2020 was considered as the outcome of this study.

106

107 **Assessment of covariates in health factors**

108 We divided the putative risk factors[8] into 5 categories: health behaviors (e.g., access
109 to exercise opportunities, insufficient sleep), clinical care (e.g. primary care
110 physicians ratio), social and economic factors (e.g., racial segregation index), physical
111 environment (e.g., transit problems and air quality), and demographics (age, sex,
112 rural, and race/ethnicity). For example, there were previous studies which identified
113 the air pollution may relate to high levels of COVID-19[10] and elder population had
114 the high risk in the COVID-19[11]. Besides these identified risk factors, we were
115 interested in the adverse health factors may link to the mortality of COVID-19. Table
116 1 presented descriptive definition, sources and literature of 12 risk factors. All deaths
117 resulted from complications of COVID-19.

118

119 **Statistical analysis**

120 The trend of the cumulative confirmed cases varied greatly in counties of the
121 United States. We used the partitioning around medoids (PAM) clustering
122 algorithm[12, 13] to assign counties with similar trends into a homogenous class after
123 standardizing the time series of cumulative confirmed cases from March 1 to August
124 27, 2020. Based on the clustering results, we used the Kruskal-Wallis test[14] to

125 detect whether there were significant differences in the distributions of 12 risk factors
126 across different classes of counties. The 12 risk factors were used to build a negative
127 binomial model[15, 16] for every class of the counties. The analysis was conducted in
128 R version 3.6.1.

129

130 **Validation analysis**

131 We randomly divided counties (samples) into training (70% of the counties) and
132 testing (30% of the counties) in each class. The model obtained from the training data
133 was employed to predict the death counts of COVID-19 in the testing data, and the
134 accuracy was assessed by the root mean square error (RMSE) of the mortality ratio
135 (the number of deaths divided by the number of cumulative confirmed cases).

136

137 **Results**

138 **Three classes of county-level infection in the United States**

139 The clustering analysis grouped the 3,125 counties were assigned into 3 classes.
140 There were 2,751 counties in the first class with the lowest overall cumulative
141 confirmed cases. Its medoid was Halifax County in Virginia. There were 294 counties
142 in the second class with a median level of overall cumulative confirmed cases. Its
143 medoid was St. Clair County in Illinois. There were 80 counties in the third class with
144 the highest overall cumulative confirmed cases. Its medoid was Marion County in
145 Indiana. Here, the PAM algorithm selected the county with most representative data

146 as the medoid in a class[12, 13]. The geographical distribution of the counties by class
147 was shown in Figure 1, where the size of a circle indicated the cumulative confirmed
148 cases on August 27, 2020. The distribution of deaths on August 27, 2020, which
149 clearly differed among the three classes, was also presented in Figure 1. Note that the
150 east, south, and west coasts were the most severely hit areas by COVID-19. Most
151 counties in the high prevalence class were from Massachusetts, New York, New
152 Jersey, Florida, Texas and California [9].

153

154 **Distributions of 12 selected risk factors in the three classes of counties**

155 Figure 2 showed the distributions of the 12 selected risk factors by the class of
156 counties. The distributions were significant different ($P<0.001$) for all 12 risk factors.
157 For example, the average population in the low prevalence class was 38,444, which
158 was 10% and 3% of the average populations in the median and high prevalence
159 classes, respectively. The average proportion of rural residents in the low prevalence
160 class was 64.47%, versus 2.72% in the high prevalence class. The segregation index
161 of non-Whites versus Whites was the largest in the high prevalence class, but the
162 smallest in the low prevalence class.

163

164 **Factors influencing mortality of COVID-19 in the three classes**

165 There were three common factors, namely, residential segregation between non-
166 Whites and Whites, resident population, and the percentage of Hispanic population,

167 which had statistically significant ($P<0.05$) effects on mortality in all classes. The
168 negative binomial model was used to understand the within-class effects of residential
169 segregation between non-Whites and Whites and the percentage of Hispanic
170 population on mortality of COVID-19 as shown in Figure 3. Note that the higher
171 values of both residential segregation between non-Whites and Whites and the
172 percentage of Hispanic population the higher mortality of COVID-19. In the high
173 prevalence class, an increase in both the residential segregation between non-Whites
174 and Whites and the percentage of Hispanic population resulted in more deaths than
175 other two classes of counties.

176 Table 2 presented the significant factors specific to each class based on the
177 training data. Specifically, in the low prevalence class, nine variables were
178 significantly associated with the mortality of COVID-19. Higher values in the
179 average daily density of PM_{2.5} (0.153, $P<0.001$), the percentage of workforce driving
180 alone to work (0.039, $P<0.001$), the percentage of workforce that had more than 30
181 minutes commute driving alone (0.015, $P<0.001$), the percentage of adults who
182 reported less than average 7 hours sleeping (0.073, $P<0.001$), resident population
183 ($P<0.001$), the percentage of Hispanic (0.020, $P<0.001$) and female population
184 (0.054, $P<0.001$) and segregation index (0.015, $P<0.001$) , significantly increased the
185 number of deaths, while more people living in rural areas (-0.014, $P<0.001$) decreased
186 the number of deaths of COVID-19.

187 In the median prevalence class, eight variables were significantly associated with
188 the deaths of COVID-19. Higher values in the percentage of workforce driving alone
189 to work ($P=0.016$, $P=0.008$), the percentage of workforce that had more than 30 minutes
190 commute driving alone ($P<0.001$), resident population ($P<0.001$), the
191 percentage of population aged over 65 ($P<0.001$), the percentage of Hispanic
192 ($P<0.020$, $P<0.001$) and female population ($P=0.127$, $P=0.001$) and segregation index
193 ($P<0.025$, $P<0.001$)-led to an increase in deaths as opposite to a decrease in deaths of
194 COVID-19 for more people living in rural areas ($P=0.011$, $P=0.007$) .

195 In the high prevalence class, four variables were significantly associated with
196 mortality. Higher values in resident population ($P<0.001$), , the percentage of
197 population aged over 65 ($P<0.001$), the percentage of Hispanic population
198 ($P=0.014$, $P=0.009$) and segregation index ($P=0.019$, $P=0.005$), caused more deaths.

199 For each class of counties, the model obtained from the training data was
200 employed to predict the deaths of COVID-19 on August 27, 2020 using the testing
201 data. The corresponding RMSE values for the mortality ratio were 0.056%, 0.041%,
202 and 0.088%, respectively, in the low, median, and high prevalence classes.

203

204 **Discussion**

205 Using the time trends of the cumulative confirmed cases in 3,125 counties in the
206 United States, we categorized those counties into three levels of infection. The low
207 prevalence class counted for 88% of the 3,125 counties. Their resident population was

208 remarkably smaller than the other two classes of counties. But the resident population
209 size increased the mortality of COVID-19 regardless of the level of COVID-19
210 prevalence. A higher population density may increase more contacts in social
211 distancing[17, 18], leading to a higher risk in mortality of COVID-19. On the
212 contrary, a higher percentage of residents living in rural areas in both the low and
213 median prevalence classes of counties may reduce the mortality. Disparities in race
214 and ethnicity were found in the infected populations. For example, Blacks were
215 reported to be prone to COVID-19[19, 20], and living settings of racial/ethnic
216 minorities were founded to be more crowded, making social distancing difficulty [21].
217 In this study, we found that Hispanics were more vulnerable. Further investigation is
218 warranted to study the racial disparity in the mortality of COVID-19. However, the
219 segregation index between non-Whites and Whites revealed the racial disparity in
220 health, leading to differences in health status not only at the individual level but also
221 at the community level[22]. A higher values in the segregation index indicated the
222 poor health status, which may increase the mortality of COVID-19[17]. This health
223 inequality increased the mortality rates of COVID-19 in all classes of counties.

224 For the low and the median prevalence class of counties, more workforce driving
225 alone to work and commuting long-distance may increase the levels of anxiety[23],
226 leading to the high mortality in COVID-19. A higher percentage of long-distance
227 commuting workforce was also linked to a high level of anxiety for commuters[23].
228 And substantial time spent by long-distance commuters could inhibit their healthy

229 behaviors[24]. The stress and less healthy behaviors may increase individual's
230 vulnerability to COVID-19[27-29]. Also, long-distance commute may be necessary
231 for people who work in relatively higher dense areas where the risk of COVID-19 is
232 high.

233 The counties in both the low and the median class of prevalence were accounted
234 for 97.44% of the infected counties, where the higher values in the percentage of
235 female population increased the mortality of COVID-19.

236 The percentage of adults with inadequate sleeping time was found to increase the
237 mortality of COVID-19 in the low prevalence class of counties. Sleeping time was
238 reported to be associated with the health system[25]. The higher number of people
239 who had inadequate sleeping time, the more adverse effects of sleep on immunity
240 were identified[26]. The air quality also was reported to be associated with the
241 mortality rate of COVID-19[10, 30, 31].

242 For both the median and the high prevalence class of counties, there was an age
243 trend in the mortality rate of COVID-19. In those counties, there was a higher
244 percentage of elderly, indicating a larger population of individuals aged over 65,
245 which increased the mortality rate of COVID-19[11].

246

247 **Conclusion**

248 This study identified several significant risk factors associated with the mortality
249 of COVID-19, and our findings are highly valuable and timely for the decision-

250 makers to develop strategies in reducing the mortality of COVID-19. The study relied
251 on mortality data on August 27, 2020. The counties were randomly divided into the
252 training and testing data once. However, we offered the epidemiological picture to
253 facilitate the identification of important factors influencing the mortality of COVID-
254 19 across different levels of infected counties in the United States. Regardless of the
255 regions, the factors linked to the poor health status contributed to higher mortality of
256 COVID-19. Improving the clinical care and eliminating the racial health inequality,
257 combined with improving physical environment were expected to significantly
258 decrease the mortality rate of COVID-19. Thus, we recommended that local
259 governments should reduce physical and psychological risks in residential
260 environments.

261

262 **Declarations**

263 **Ethics approval and consent to participate**

264 Not applicable

265 **Consent for publication**

266 Not applicable

267 **Availability of data and materials**

268 The data that support the findings of this study are available from the New York

269 Times and the County Health Rankings and Roadmaps program website. The data

270 and R files supporting the conclusions of this article are available in the

271 <https://github.com/tingT0929/Risk-factors-associated-with-mortality-of-COVID-19>.

272 **Competing interests**

273 The authors declare that they have no competing interests

274 **Funding**

275 Not applicable

276 **Authors' contributions**

277 TT, XW and HZ developed the idea and research. TT, JZ, LH and YJ wrote the first

278 draft of the manuscript and all other authors discussed results and edited the

279 manuscript. JZ, LH, and YJ collected and validated epidemiological data. CD and ZL

280 collected and compiled census data. All authors read and approved the final

281 manuscript

282 **Acknowledgements**

283 We would like to thank all individuals who are collecting epidemiological data of the

284 COVID-19 outbreak around the world.

285

286 **References**

- 287 1. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich
288 NG, Lessler J: **The Incubation Period of Coronavirus Disease 2019 (COVID-19)**
289 **From Publicly Reported Confirmed Cases: Estimation and Application.** *Annals*

- 290 *of Internal Medicine* 2020.
- 291 2. **Distribution of COVID-19 cases in the world** [<http://2019ncov.chinacdc.cn/2019->
292 [nCoV/global.html](http://2019ncov.chinacdc.cn/2019-ncov/global.html)]
- 293 3. **Coronavirus Disease 2019 (COVID-19)** [<https://www.cdc.gov/coronavirus/2019->
294 [ncov/cases-updates/cases-in-us.html](https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html)]
- 295 4. Hadler JL, Yousey-Hindes K, Pérez A, Anderson EJ, Bargsten M, Bohm SR, Hill M,
296 Hogan B, Laidler M, Lindegren ML: **Influenza-related hospitalizations and**
297 **poverty levels—United States, 2010–2012.** *Morbidity Mortality Weekly Report*
298 2016, **65**(5):101-105.
- 299 5. Noppert GA, Yang Z, Clarke P, Ye W, Davidson P, Wilson ML: **Individual-and**
300 **neighborhood-level contextual factors are associated with Mycobacterium**
301 **tuberculosis transmission: genotypic clustering of cases in Michigan, 2004–2012.**
302 *Annals of epidemiology* 2017, **27**(6):371-376. e375.
- 303 6. Maier HE, Lopez R, Sanchez N, Ng S, Gresh L, Ojeda S, Burger-Calderon R, Kuan
304 G, Harris E, Balmaseda A: **Obesity increases the duration of influenza a virus**
305 **shedding in adults.** *The Journal of infectious diseases* 2018, **218**(9):1378-1382.
- 306 7. Lin K, Fong DY-T, Zhu B, Karlberg J: **Environmental factors on the SARS**
307 **epidemic: air temperature, passage of time and multiplicative effect of hospital**
308 **infection.** *Epidemiology & Infection* 2006, **134**(2):223-230.
- 309 8. **State Rankings Data & Reports**
310 [<https://www.countyhealthrankings.org/reports/county-health-rankings-reports>]

- 311 9. **Coronavirus in the U.S.: Latest Map and Case Count**
- 312 [\[https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html\]](https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html)
- 313 10. Conticini E, Frediani B, Caro D: **Can atmospheric pollution be considered a co-**
- 314 **factor in extremely high level of SARS-CoV-2 lethality in Northern Italy?**
- 315 *Environmental pollution* 2020:114465.
- 316 11. Onder G, Rezza G, Brusaferro S: **Case-Fatality Rate and Characteristics of**
- 317 **Patients Dying in Relation to COVID-19 in Italy.** *JAMA* 2020.
- 318 12. Zhang LS, Yang MJ, Lei DJ: **An improved PAM clustering algorithm based on**
- 319 **initial clustering centers.** In: *Applied Mechanics and Materials: 2012*: Trans Tech
- 320 Publ; 2012: 244-249.
- 321 13. Lei D, Zhu Q, Chen J, Lin H: **Automatic PAM Clustering Algorithm for Outlier**
- 322 **Detection.** *Journal of Software* 2012.
- 323 14. Brunner E, Konietzschke F, Bathke AC, Pauly M: **Ranks and Pseudo-Ranks-**
- 324 **Paradoxical Results of Rank Tests.** *arXiv preprint arXiv:180205650* 2018.
- 325 15. Hilbe JM: **Negative binomial regression:** Cambridge University Press; 2011.
- 326 16. Zeileis A, Kleiber C, Jackman S: **Regression models for count data in R.** *Journal of*
- 327 *statistical software* 2008, **27**(8):1-25.
- 328 17. Dowd JB, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, Liu Y, Mills MC:
- 329 **Demographic science aids in understanding the spread and fatality rates of**
- 330 **COVID-19.** *Proceedings of the National Academy of Sciences* 2020.
- 331 18. Greenstone M, Nigam V: **Does social distancing matter?** *University of Chicago,*

- 332 *Becker Friedman Institute for Economics Working Paper* 2020(2020-26).
- 333 19. Hooper MW, Nápoles AM, Pérez-Stable E: **COVID-19 and Racial/Ethnic**
- 334 **Disparities.** *JAMA* 2020.
- 335 20. Laurencin CT, McClinton AI: **The COVID-19 pandemic: a call to action to identify**
- 336 **and address racial and ethnic disparities.** *Journal of Racial Ethnic Health*
- 337 *Disparities* 2020:1-5.
- 338 21. Noonan AS, Velasco-Mondragon HE, Wagner FAJPhr: **Improving the health of**
- 339 **African Americans in the USA: an overdue opportunity for social justice.** 2016,
- 340 **37(1):1-20.**
- 341 22. Williams DR, Collins C: **Racial residential segregation.** *Race, Ethnicity, Health: A*
- 342 *Public Health Reader.*2012, **26**:331.
- 343 23. Van Rooy DL: **Effects of automobile commute characteristics on affect and job**
- 344 **candidate evaluations: A field experiment.** *Environment Behavior* 2006, **38(5):626-**
- 345 **655.**
- 346 24. Wolin KY, Bennett GG, McNeill LH, Sorensen G, Emmons KMJAjohb: **Low**
- 347 **discretionary time as a barrier to physical activity and intervention uptake.** 2008,
- 348 **32(6):563-569.**
- 349 25. Besedovsky L, Lange T, Haack M: **The sleep-immune crosstalk in health and**
- 350 **disease.** *Physiological reviews* 2019, **99(3):1325-1380.**
- 351 26. Irwin M: **Effects of sleep and sleep loss on immunity and cytokines.** *Brain,*
- 352 *behavior, immunity* 2002, **16(5):503-512.**

- 353 27. Mazza C, Ricci E, Biondi S, Colasanti M, Ferracuti S, Napoli C, P. R: **A Nationwide**
354 **Survey of Psychological Distress among Italian People during the COVID-19**
355 **Pandemic: Immediate Psychological Responses and Associated Factors.**
356 *International Journal of Environmental Research and Public Health* 2020,
357 **17(9):3165.**
- 358 28. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y: **A nationwide survey of**
359 **psychological distress among Chinese people in the COVID-19 epidemic:**
360 **implications and policy recommendations.** *Gen Psychiatr* 2020, **33(2):e100213-**
361 **e100213.**
- 362 29. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, Ho RC: **Immediate psychological**
363 **responses and associated factors during the initial stage of the 2019 coronavirus**
364 **disease (COVID-19) epidemic among the general population in China.**
365 *International journal of environmental research public health* 2020, **17(5):1729.**
- 366 30. Wu X, Nethery RC, Sabath BM, Braun D, Dominici F: **Exposure to air pollution**
367 **and COVID-19 mortality in the United States.** *medRxiv* 2020.
- 368 31. Contini D, Costabile F: **Does Air Pollution Influence COVID-19 Outbreaks?** In.:
369 Multidisciplinary Digital Publishing Institute; 2020.
- 370 32. Martelletti L, Martelletti P: **Air pollution and the novel Covid-19 disease: a**
371 **putative disease risk factor.** *SN Comprehensive Clinical Medicine* 2020:1-5.
- 372 33. Wenham C, Smith J, Morgan R: **COVID-19: the gendered impacts of the outbreak.**
373 *The Lancet* 2020, **395(10227):846-848.**

- 374 34. Abdulmir AS, Hafidh RR: **The Possible Immunological Pathways for the**
375 **Variable Immunopathogenesis of COVID--19 Infections among Healthy Adults,**
376 **Elderly and Children.** *Electronic Journal of General Medicine* 2020, **17**(4).
377

Table 1. The definitions and sources of the 12 selected risk factors, and the literature supporting these factors.

Category	Variable	Description	Source	Related Topics	Previous Studies
HEALTH BEHAVIORS	% With Access to Exercise Opportunities	Percentage of population with adequate access to locations for physical activity.	Business Analyst, Delorme map data, ESRI, & US Census Tigerline Files	Social Distancing; Outdoor and indoor exercise	Novel
	% Insufficient Sleep	Percentage of adults who report fewer than 7 hours of sleep on average.	Behavioral Risk Factor Surveillance System	Immune system	Novel
CLINICAL CARE	Primary Care Physicians Ratio	Ratio of population to primary care physicians.	Area Health Resource File/American Medical Association	Medicare services	Novel
SOCIAL & ECONOMIC FACTORS	Segregation index-non-Whites/Whites	Index of dissimilarity where higher values indicate greater residential segregation between non-White and White county residents.	American Community Survey, 5-year estimates	Poor health status	Novel
PHYSICAL ENVIRONMENT	% Long Commute - Drives Alone	Among workers who commute in their car alone, the percentage that commute more than 30 minutes.	American Community Survey, 5-year estimates	Psychological risks	Novel
	% Drive Alone to Work	Percentage of the workforce that drives alone to work.	American Community Survey, 5-year estimates	Psychological risks	Novel

	Average Daily PM2.5	Average daily density of fine particulate matter in micrograms per cubic meter (PM2.5).	Environmental Public Health Tracking Network	Air quality	E Conticini, B Frediani and D Caro [10] D Contini and F Costabile [31] L Martelletti and P Martelletti [32]
DEMOGRAPHICS	Population	Resident population.	Census Population Estimates	Population densities	Novel
	% Rural	Percentage of population living in a rural area.	Census Population Estimates	Demographical science	Novel
	% Hispanic	Percentage of population that is Hispanic.	Census Population Estimates	Minorities	Novel
	% Female	Percentage of population that is female.	Census Population Estimates	Sex	C Wenham, J Smith and R Morgan [33]
	% 65 and over	Percentage of population ages 65 and older.	Census Population Estimates	Age structure	G Onder, G Rezza and S Brusaferrò [11], [34]

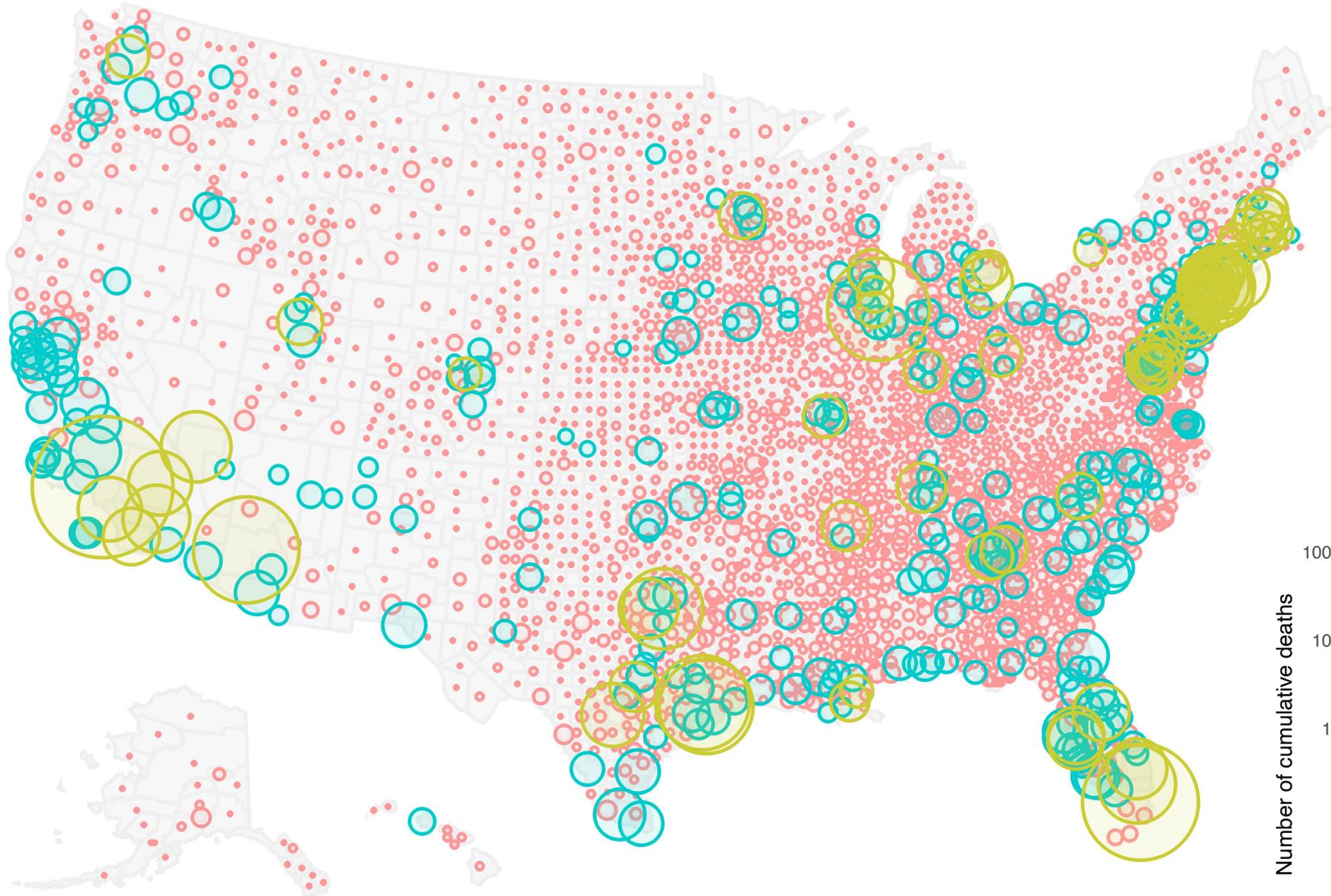
Table 2. Variables significantly related to the mortality rate of COVID-19 in the three classes of counties.

	Variable	Estimate	Pr(> t)
Low prevalence class	`Average Daily PM2.5`	0.153	<0.001
	`% Drive Alone to Work`	0.039	<0.001
	`% Long Commute - Drives Alone`	0.015	<0.001
	`% Insufficient Sleep`	0.073	<0.001
	Population	1.00E-05	<0.001
	`% Hispanic`	0.020	<0.001
	`% Female`	0.054	<0.001
	`% Rural`	-0.014	<0.001
	`Segregation index-non-White/White`	0.015	<0.001
Median prevalence class	`% Drive Alone to Work`	0.016	0.008
	`% Long Commute - Drives Alone`	0.013	<0.001
	Population	1.07E-06	<0.001
	`% 65 and over`	0.049	<0.001
	`% Hispanic`	0.020	<0.001
	% Female	0.127	0.001
	`% Rural`	-0.011	0.007
	`Segregation index-non-White/White`	0.025	<0.001
High prevalence class	Population	2.22E-07	<0.001
	`% 65 and over`	0.114	<0.001
	`% Hispanic`	0.014	0.009
	`Segregation index-non-White/White`	0.019	0.005

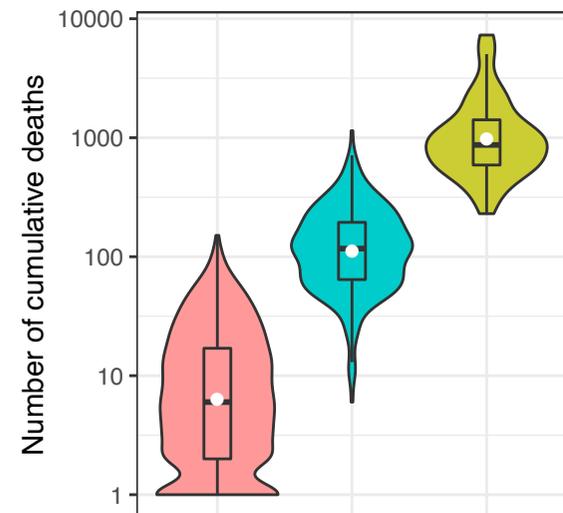
Figure 1. The geographical distribution of three classes of counties. The clustering was based on time-variant cumulative confirmed cases from March 1 to August 27, 2020. The size of circle represented the total confirmed cases on August 27, 2020. The distributions of deaths on August 27, 2020 in the three classes of counties were combined.

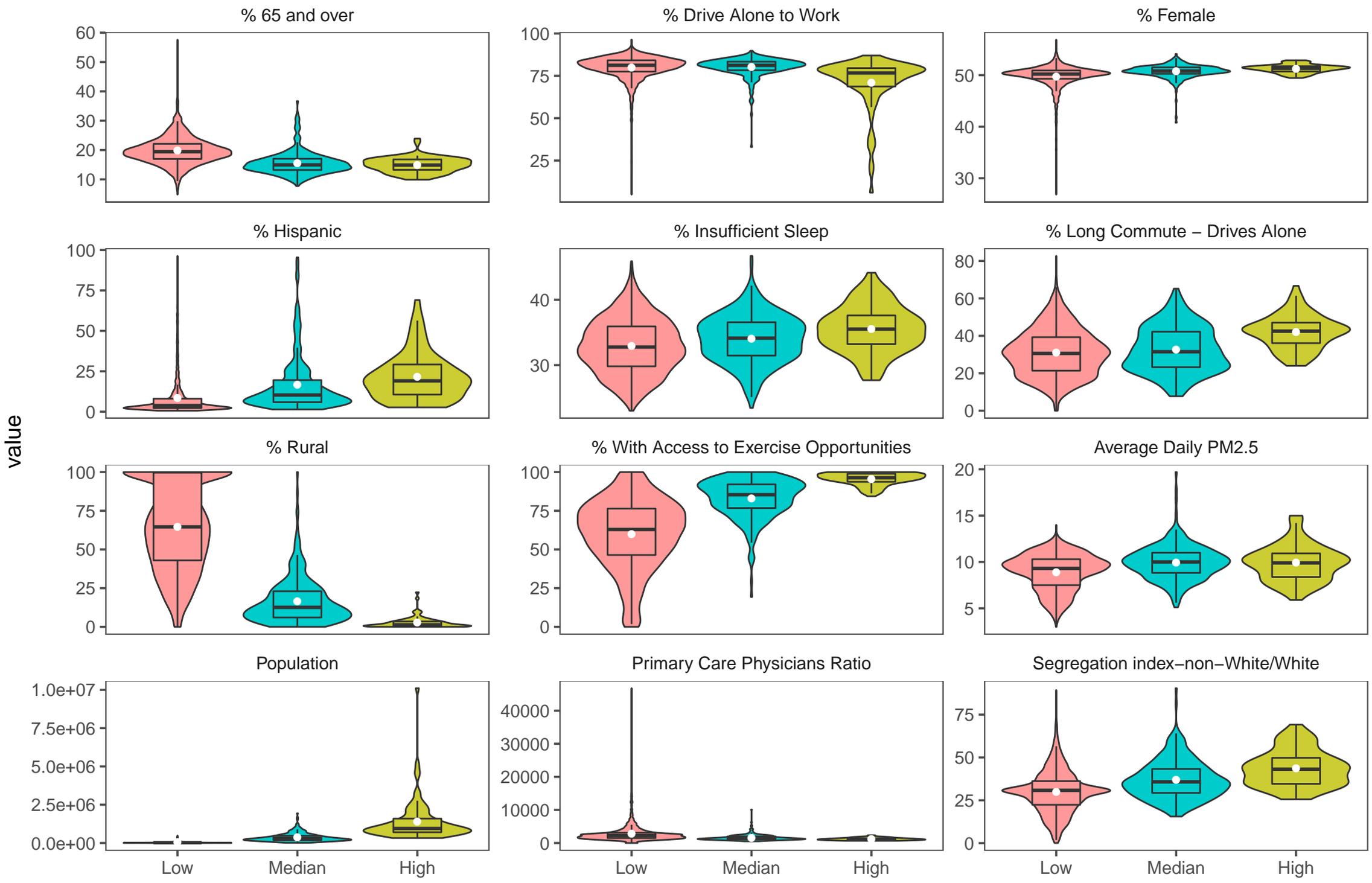
Figure 2. Violin diagram and boxplot of the distributions of the 12 selected risk factors in the three classes of counties. The low, median and high prevalence classes of counties were represented by red, blue and green colors.

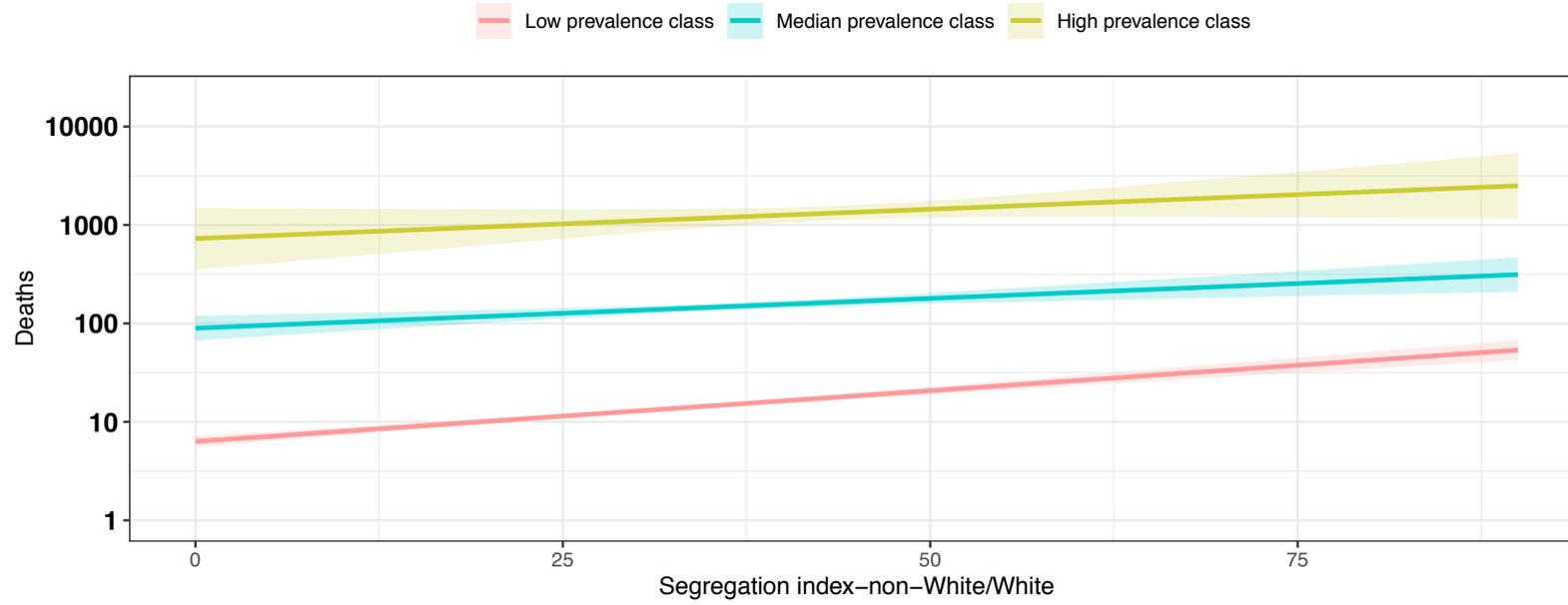
Figure 3. The risk factors common in the three classes of counties. A: Segregation index for non-Whites versus Whites; B: The percentage of Hispanic population. Direct curves were generated using the negative binomial model with the segregation index between non-Whites and Whites or the percentage of Hispanic population as the single covariate.



○ Low prevalence class ○ Median prevalence class ○ High prevalence class





A**B**