Estimating the Size of High-risk Populations for COVID-19 Mortality across 442 US Cities

Jin Jin[1,*], Neha Agarwala[2,*], Prosenjit Kundu[1,*], Nilanjan Chatterjee[1,3,**]

[1]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University
[2]Department of Mathematics and Statistics, University of Maryland, Baltimore County
[3]Department of Oncology, School of Medicine, Johns Hopkins University

[*]These authors contributed equally
[**]Correspondence to: nilanjan@jhu.edu

Abstract

A variety of predisposing factors have been associated with serious illness and death from COVID-19. Understanding the distribution of risks associated with these factors by local communities can provide important opportunities for targeting interventions. We characterize the distribution of risk for COVID-19 mortality for populations at large across 442 US cities, by utilizing recently published estimates of risk associated with age, gender, ethnicity, social deprivation and 12 health conditions from a very large UK-based study, combined with the information available on prevalence and co-occurrence of these factors in the US through a variety of population-based public databases. We estimate that across all the cities, an underlying weighted risk-score can identify a total of approximately 12.65 million, 4.09 million and 1.34 million individuals who are at 2-, 5- and 10-fold higher risk, respectively, compared to the average risk for the US population. The percentage of population which exceed the respective risk thresholds varies across the cities in the range (1st-99th percentile), 3.6%-20.1%, 0.7%-8.0% and 0.1%-3.2%, respectively. The percentage of deaths within a city that are expected to occur above these risk-thresholds varies in the range of 20.1%-53.5%, 8.5%-38.2% and 2.9%-25.4%, respectively. Our analysis can provide guidance to national and local policy makers regarding resources needed to protect the most vulnerable populations in these communities, and how much utility such interventions may have in reducing the total population burden of death.

Introduction

The first case of SARS-CoV-2 infection in the US was reported on January 20th, 2020, in the state of Washington [1,2], and to date the pandemic has led to nearly 100,000 COVID-19 deaths - making US by far the most affected country globally. There is, however, major variation in rates of infections and underlying deaths across US states, counties and cities. Various local population characteristics, such as mitigation measures [3,4], population density and mobility patterns [5,6] define background risks of illness and death across the regions. Further, epidemiologic studies [7-16] are providing evidence for pre-disposing factors that can put individuals at differential risks of serious illness and mortality.

In the US, both the number of reported daily infections and the number of reported daily deaths have recently reached peak, but the post-peak decline of these numbers has been slow [17]. During the first phase of the pandemic, US and other countries have relied on broad and strict intervention measures, such as country/state-wide lockdowns and travel restrictions. However, as it becomes evident that the pandemic is likely to last for months and possibly years to come, mitigation efforts in the future will rely on both broad but more relaxed measures, such as social distancing, and more strict intervention for targeting towards high-risk populations and individuals. Clearly, a large fraction of deaths has occurred among individuals of old age, and in US and other western countries, community living in nursing home settings has been a major source of risk for these individuals. Further, serious illness and death have been shown to be more common among male, various minority populations, and individuals with selected health conditions [10,14]. As lockdown and travel restrictions are lifted, measures will need to stay in place to protect these high-risk individuals through "shielding" [18] and prioritization for scarce preventive resources [19,20]. As future planning for such effort requires understanding the size of "high-risk" populations, a few studies have now emerged to provide such information for UK [21], US [22] and globally by nations and regions [23]. All of these studies, however, define high-risk group in a broad fashion based on risk-factor prevalence, without specific definition of the level of the underlying risk.

In this article we report results from our study for estimating the size of general populations who are at various levels of risk for COVID-19 mortality due to predisposing factor across a large

number of US cities. We use recently published results from a large UK-based study on risk of mortality associated with a variety of predisposing factors, which could influence risk of infection or fatality or both [14]. We define a risk-score based on multivariate adjusted risk estimates and combine it with information on prevalence and co-occurrence of these factors from data sources available from various national agencies. We use a series of novel methods to obtain estimates of proportion of individuals within each city who exceed different risk-thresholds. We also provide projections for the number of deaths that are expected to arise within the defined high-risk groups, as a percentage of the total number of deaths in the underlying city populations.

Results

We observe wide variation in the underlying risk-score values across individuals who participated in National Health Interview Survey (Supplementary Figure 1). The value of the risk-score at the 99th and 1st percentile of the distribution corresponds to a risk ratio of approximately 8-fold among the age-group 18-39, and 305-fold among the age group 40+. Overall, we observe that 12.3%, 4.4% and 1.4% of individuals are at or above risk-thresholds associated with elevated (>2-fold), high (>5-fold) and very-high risk (>10-fold) categories (Table 1). A small, but not negligible, fraction of the population exceeds the threshold for extremely high-risk (25-fold). The percentage of the populations exceeding these thresholds vary strongly by age. In particular, only a small fraction (<3%) of individuals who are younger than 70 exceed the threshold for high-risk. In contrast, majority of the people who are 80 years or older are at high-risk, and a quarter of them are at very high-risk. We further examine the distribution of various other risk factors among individuals in the defined high-risk groups (Supplementary Figures 2-3). As expected from the nature of risk-factor association, male, Hispanic and African Americans, and individuals with obesity and various health conditions are more common in the different risk groups compared to the general NHIS population. In addition, some factors, such as former smoking and hypertension, which were not identified to be strong risk-factors in the UK study, appear to be more prevalent in the high-risk groups because of their association with strong risk-factors, such as age and type-2 diabetes.

We observe substantial variation in prevalence of all the risk-factors across the different cities (Supplementary Figures 4-5). We evaluate the Index of Excess Risk for COVID-19 (IER-C19) mortality, a measure of aggregated risk in a locality associated with the prevalence of the underlying factors (Figure 1). We observe an almost 8-fold risk ratio between the highest and the lowest ranked city according to this index. Five cities with the highest values of the index are Hemet (CA), Detroit (MI), Youngstown (OH), Shreveport (LA) and Deerfield Beach (FL), and five cities with the lowest values of the index are Provo (UT), Frisco (TX), Fishers (IN), West Jordan (UT) and Allen (TX). A number of large cities in the US east coast tend to rank high in this index. Notably, Detroit and Ann Arbor, two cities only separated by 43.4 miles within Michigan, rank in the two extremes with almost 4-fold difference in average risk.

The proportion of individuals in different risk categories varies widely across cities (Figure 2, Supplementary Figure 6, Supplementary Table 1). Our analysis identifies 93 cities which have at least 5% of individuals at high-risk category (>5-fold risk). These same cities have at least 14.2% and 1.6% of individuals who could be classified at elevated (>2-fold) and very-high (>10-fold) risk groups. The number of individuals in the high-risk categories depends heavily on the underlying population sizes. In the New York City, we estimate that there are 1228K, 413K and 139K individuals who exceed these risk-thresholds for elevated, high and very high-risk categories. A number of other major population centers, including Los Angeles, Chicago, Philadelphia, Houston, San Antonio and Detroit, have at least 64K people, in each city, who can be classified at the >5-fold risk category. The number of individuals who are in the elevated- and very- high-risk categories across these cities varies in the range of 150K-517K and 27K-53K, respectively.

Our analysis also identifies 78 cities which have less than 2% of individuals in the high-risk category (>5-fold). A number of these cities are known for large University associated populations, including Cambridge (MA), College Station (TX), Iowa City (IA), Ann Arbor (MI) and Madison (WI). Other notable cities included in the list are Boulder (CO), Columbia (MO), Fort Collins (CO), Provo (UT) and Salt Lake City (UT). In this list, the large cities that have population size >400K include Seattle (WA), Austin (TX), Raleigh (NC), Colorado Springs (CO) and San Jose (CA). The estimated number of individuals in these cities who exceed the risk-thresholds for the elevated, high and very high-risk categories ranges between 29K-79K, 7K-19K and 2K-5K, respectively.

The projected number of deaths that are expected to arise at or above various risk thresholds shows how deaths in the underlying populations arise disproportionately from a relatively small fraction of the population (Figure 3, Supplementary Figure 7, Supplementary Table 2). There are 94 cities, including major cities like New York, where we estimate that more than 25% of the total deaths are expected to have risen from a relatively small fraction (<5%) of high-risk individuals. We estimate that in the New York City, 43%, 27% and 16% of the deaths occurred within 14.2%, 4.8%, and 1.6% of the populations at the highest risk. Based on estimates of the total number of excess deaths due to COVID-19 in NYC until May 2 [24], we project that the absolute numbers of deaths attributable to these high-risk categories are 10358, 6637 and 3859, respectively.

Discussion

In this article, we have characterized distribution of risk associated with a set of predisposing factors for COVID-19 death across a large number of US cities. We have utilized information on recently published estimate of risk of mortality associated with these factors from a large UK study [14], prevalence of the same factors from multiple population-based data sources, individual-level data available on a nationally representative study, and novel statistical methods to estimate size of populations exceeding precisely defined risk-thresholds. Our results identify cities, including major metropolitan hubs, that have concentration of high-risk individuals. These results can provide guidance to local and national agencies for planning more targeted intervention efforts for high-risk individuals.

Mitigation efforts for the pandemic in most countries to date have focused on broad and strict intervention measures through series of lockdowns and travel restrictions. Additional efforts for targeting high-risk individuals have been generally limited. In England, about 1.5 million individuals who are at extremely high risk due to selected conditions were identified based on national health records, and were provided with government assistance for food delivery and medicine services [21]. In California, local and state government developed the Project Roomkey [25] to provide free hotel room, meal and other services to asymptomatic homeless people who are at

high risk due to their age or/and underlying conditions. In the future, as the statewide lockdowns are lifted, more initiatives for shielding high-risk individuals, starting with those who may be particularly susceptible to exposures, such as front-line workers and older population living in community settings, will be needed.

A few recent studies have investigated the proportions of "high-risk" individuals for COVID-19 related serious illness or mortality in the UK, US and across nations globally [21-23]. Further, the New York Times has recently produced a county-level map for the US to describe prevalence of some of these risk-factors [26]. These studies have defined high-risk individuals based on prevalence of one or more risk-factors, without taking into account the relative contribution of these factors. Further, because of the broad definition used, they estimate that a very large fraction of populations, 20% in UK and 16-31% across nations globally, are at "high-risk". In contrast, we have defined different risk-categories based on an underlying score that allows one to assign more precise magnitude of risks to these categories. As a result, we have been able to show that it is possible to identify smaller groups of high-risk individuals which lead to disproportionately large number of deaths across different US cities. Efforts for any targeted interventions, such as government assistance for "shielding", may not be economically viable if the definition of high-risk group becomes too broad.

Our analysis also shows that a large fraction of total deaths will occur outside of small high-risk groups. In NYC, for example, we estimate that 43% and 27% of deaths are expected to arise from 14.2% and 4.8% of the population who are at the highest risk. The estimate implies that a majority of deaths will occur outside of these risk groups. In particular, we observe that the current set of risk-factors have very limited ability to identify individuals who are younger than 60 at high risk groups (see Table 1) and yet current data suggest that a substantial fraction of deaths will arise from such younger age groups. Thus, targeted intervention for elevated and high-risk individuals through shielding and other efforts, cannot be a substitute for broader community level intervention through social distancing and other measures. Further, research is urgently needed for identifying additional risk-factors, including genetic predisposition and other biomarkers, which can better identify younger individuals who are likely to face serious illness and mortality.

In this article, we investigate the potential excess risks faced by cities, and individuals within cities, due to various predisposing factors. The absolute risks of these communities and individuals, however, heavily depend on the underlying local characteristics of the epidemic driven by key factors such as population density, mobility patterns and social distancing. Estimates available based on excess death, for example, indicate a mortality rate for the NYC from COVID-19 about 283 per 100K individuals during the period of March 13-May 2 [27,28]. According to our estimate, the rate of death in the high-risk group (>5-fold) is expected to be about 1620 per 100K individuals. Now, consider a hypothetical scenario where the pandemic returns with double its intensity later this year. Thus, over a similar period of time, such a resurgence will lead to a death rate due to COVID-19 of 566 and 3240 per 100K individuals, in the overall city and in the high-risk group, respectively. The increase in absolute risk due to doubling the intensity of the pandemic in these two groups will be 283 vs 1620 per 100K individuals, indicating a much more adverse impact on the individuals in the high-risk group. In general, our framework can be used to model absolute risk of different risk-groups under various types of pandemic scenarios typically evaluated by the forecasting models [29].

While we present the most sophisticated analysis of its kind, our study has several limitations as well. We lacked individual-level data at the level of cities and thus proposed a series of approximations to estimate the distribution of risk. We estimate co-occurrence rates of various risk-factors based on underlying prevalence and odds-ratio measures of aggregation estimated from the nationally representative NHIS. Further, we use the individual-level data available from the NHIS study to evaluate the accuracy of the mixture normal approximation for estimating the proportion of high-risk individuals (Supplementary Figure 1). In the future, accuracy of the approximation may be further improved by using alternative distributional assumptions.

We assumed that the degree of association of COVID-19 death with various predisposing factors observed in the large UK study will be generalizable to the US population [30-33]. While a number of US-based studies [12,27] using case series have reported overrepresentation of many of these factors among patients with severe illness, no large scale population-based epidemiologic studies are available to report precise risk associated with these factors in the US setting. In general,

relative-risks associated with major predisposing factors for various outcomes, including communicable [33,34] and non-communicable diseases [32], tend to be similar between US and UK. The New York City Health Department publishes population-based estimates of rate of hospitalization and death by age, gender and ethnic groups [35]. We found that the crude (unadjusted) rate ratios for deaths reported in NYC with these factors are fairly consistent with those reported in the UK study. In our analysis, we consider a risk-score defined by the predisposing factors with weights obtained from the fully adjusted model published by the UK study. The risk-score, however, does not consider potential interactions between various predisposing factors and thus may over-/under-estimate risk for certain combination of these factors. In the future, as results from more complex models that include additional risk-factors and their interactions become available, our estimates can be further refined within the framework we have defined.

The Ethnic characteristics of the UK and US population are substantially different. We observed that the crude ratio of COVID-19 death rate for blacks compared to whites in UK is very similar to that observed for the African American population compared to non-Hispanic whites within the NYC. The UK study further reports an increased risk for Asians or British Asians. In contrast, in NYC, the Asian population appears to be at a comparable risk as non-Hispanic whites. The difference is likely to be due to different countries of origin and socioeconomic conditions for these groups across the two countries. In our analysis, we assigned the risk of Asian in the US population to be the same as that of non-Hispanic whites. For the Hispanic population, which is absent in the UK, we obtained age-adjusted rate ratio for death compare to non-Hispanic whites based on data available from the NYC [36,37], and included an additional component of risk due to Hispanic origin. We could not find comparable risk estimates for other minority populations such as American Indians, Asian Indians and mixed races, and thus could not include a component of risk due to such ethnic origins. Nevertheless, it is likely that other predisposing conditions, such as age, gender and various health conditions will have similar link with risk of death in these populations.

The UK study reported a strong gradient of risk of COVID-19 death associated with the Index of Multiple Deprivation (IMD), an area-level measure of social deprivation. The study noted that the

association of COVID-19 death with IMD remains strong (a risk ratio of 1.70 between 5th vs 1st quartile) even after adjusting for ethnicity and the known comorbidity conditions. In our analysis, we used an alternative county-level measure of Social Deprivation Index (SDI) that is available in the US setting and assigned each US city with the SDI measure of the corresponding county to which the city belongs. We assigned the same degree of risk across the different quintiles of SDI as those observed for IMD in the UK study. Both IMD and SDI capture the same major components of deprivation, namely income, education, employment and housing conditions. Some of these characteristics are known to confer similar risks across UK and US for broad health outcomes such as disability adjusted life years [38]. The two variables, however, have some unique components, such as ownership of cars in SDI, that may lead to additional risks for COVID-19 death. Future population-based epidemiologic studies are urgently needed to characterize risk of COVID-19 infection, serious COVID-19 illness and mortality in relationship to various ethnic groups and social deprivation in the US setting.

In summary, in spite of some limitations, we present a very comprehensive and rigorous analysis of distribution of risk for COVID-19 death across large number of US cities. While these projections can be further refined as better model and data become available in the future, the current results can provide guidance to national and local policy makers regarding size of high-risk populations who may benefit most with more targeted intervention efforts. In addition, the novel methodological framework we develop and the open-source code we make available will allow similar rigorous analysis of risk across other countries using relevant datasets.

Methods

Definition of COVID-19 risk-score

The risk-score for an individual is defined as a weighted combination of various socio-demographic characteristics and predisposing health conditions, with weights defined by the relative magnitude of the contribution of these factors to the risk of death due to COVID-19. We define the risk-score primarily using information from a very large UK-based studies involving a

population of >17 million individuals among whom more than 5000 COVID-19 deaths were reported [14]. The risk-factors included age, gender, ethnicity, an area-wide measure of social deprivation and 12 different health conditions. We define the COVID-19 death risk-score for an individual as $RS_i = \sum_{k=1}^{K} \beta_k X_{ik}$, where X_{ik}'s denote binary variables indicating the categories the ith individual belongs to across different risk-factors. We use information available from Table A1 of the paper from the UK study [14] to define the level of different risk-factors and extract the corresponding log-hazard ratio values from the fully adjusted model to define the weights. We, however, adjust the risk-score to account for different ethnic composition of the US and UK populations and account for a component of risk for Hispanic population using information on age-adjusted mortality rate available from the NYC [36,37]. We note that in this definition, the "risk of mortality" refers to that of the general population, and not among infected population. Thus, the predisposing factors can increase risk of COVID-19 death due to their effect on rate of infection or/and rate of death among infected individuals. More details on definition of the risk-score can be found in Section 1 of the Supplementary Notes.

Data sources for obtaining prevalence and joint distribution of the risk-factors

US Census Bureau - American Community Survey:

American Community Survey (ACS) is a yearly basis survey that collects information on demographic, social, economic, and housing topics throughout the United States and Puerto Rico [39]. We obtain the prevalence of demographic variables across cities. Specifically, we extract information on age and gender from the 2017 table [40], and the latest information available on ethnicity from the 2018 table [41] .

Behavioral Risk Factor Surveillance System (BRFSS):

The Center for Disease Control, US, has developed the BRFSS for conducting telephone survey to collect data on various heath related factors for US residents across states, cities, and Metropolitan/Micropolitan Areas. We use the BRFSS "500 Cities: Local Data for Better Health, 2019 release" [42] to extract the prevalence on behavioral risk factors including obesity, smoking status, high blood pressure, and chronic health indicators including diabetes, asthma, chronic heart

disease, stroke/dementia, kidney disease, rheumatoid/ lupus/ psoriasis. The 2019 release is based on the 2017 questionnaire data.

United States Cancer Statistics:

The statistics are based on data collected from different cancer registries by the Centers for Disease Control Prevention (CDC) and the National Cancer Institute (NCI) [43]. We use the 2012-2016 data to obtain 5-year incidence rates at county level and overall 5-year survival rates for different cancer sites. In our study, the cancer site-specific prevalence is calculated from the incidence rate after adjusting for the survival rate. We assume the cancer prevalence in each city to be the same as that of the corresponding county to which the city belongs.

Robert Graham Center and American Community Survey:

As a proxy for the Index of Multiple Deprivation (IMD) used in the UK study, we consider an analogous measure, Social Deprivation Index (SDI), used in the US setting. SDI is an area wide measure of 7 demographic characteristics, including the indicators for less than 12 years schooling, crowding, no car, non-employed, poverty, renter occupied, and single-parent family. The measure is derived by Robert Graham Center using 5-year estimates based on 2011-2015 data from the American Community Survey (ACS) [44].

National Health Interview Survey (NHIS):

We accessed individual-level data from the NHIS of CDC. The study collects yearly cross-sectional questionnaire-based information on various health related factors for representative population of the United States [45]. We extracted risk factor information on about 20,000 adults from the 2017 NHIS data. All of the required variables, except SDI, were available for individuals in NHIS. We use the NHIS data to investigate the distribution of risk-score (excluding SDI) across the general US population, estimate co-occurrence of pairs of factors using the underlying odds-ratio parameters, and evaluate accuracy of mixture normal approximation for risk-score distribution.

Statistical Models and Methods

Similar to the UK study, in our analysis, we assume that the risk of COVID-19 death at time t for an individual i residing in location l, e.g. a city, can be described by the proportional risk model

$$\lambda_{il}(X,t) = \lambda_l(t)\exp(\sum_{k=1}^{K}\beta_k X_{ik}) = \lambda_l(t)R_{il}(\beta),$$

where $\lambda_l(t)$ denotes the baseline risk for location l due to underlying pandemic characteristics. Here t refers to calendar time since some landmark, such as the day when cumulative death reaches some minimum threshold. The average risk of the population at location l can be defined as

$$\lambda_l^A(t) = \lambda_l(t)E_l\{\exp\,(\sum_{k=1}^{K}\beta_k X_{ik})\},$$

where E_l denotes the expectation (average) with respect to distribution of the risk factors in location l.

We define the quantity $R_l(\beta)$= $E_l\{\exp(\sum_{k=1}^{K}\beta_k X_{ik})\}$ an Index of Excess Risk (IER). If two locations have same baseline rate of deaths, then the ratio of this index across them will correspond to their rate ratio associated with death and a value of IER>1 will correspond to excess death due to difference in risk-factor distribution across the two places. In our analysis, we present the scaled version of IER as $R_l(\beta)/\bar{R}$, where $\bar{R}$ denotes the weighted average of $R_l(\beta)$ across cities with population sizes as the weights. Further, we examine the distribution of $R_{il}(\beta)$ across individuals within a location to identify size of the underlying most "vulnerable" populations. For these evaluations, ideally one would require individual-level data for the set of risk-factors $X = \{X_1, \ldots\ldots X_K\}$ for a representative sample of individuals from each city. However, in the absence of such data, we develop a framework to approximate the distributions using city-specific information on prevalence, and individual-level data from a representative sample of the whole US population available from the NHIS study. Specifically, we use data from NHIS to estimate degree of co-occurrence of the different risk-factors and to evaluate the accuracy of mixture normal approximation for tail probability calculations (see Supplementary Figure 1). Further details of the methods can be found in Section 2 of the Supplementary Notes.

Code Availability

All codes for data management and the analyses in this article can be accessed at https://github.com/nchatterjeelab/COVID19Risk.

Acknowledgements

We thank Dr. Allison Meisner from the Johns Hopkins University, Biostatistics Department and Dr. Montserrat García-Closas, Division of Cancer Epidemiology and Genetics at National Cancer Center for their comments on a previous version of the manuscript. The funding for this research came from the Bloomberg Distinguished Professorship endowment.

References

1. Link A, Hold G. First Case of Covid-19 in the United States. N Engl J Med 2020;382:e53.
2. Centers for Disease Control and Prevention. First travel-related case of 2019 novel Coronavirus detected in United States. 2020. (Accessed May 27, 2020, at https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html.)
3. Giordano G, Blanchini F, Bruno R, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020:1-6.
4. Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020;323:1915-23.
5. Chinazzi M, Davis JT, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020;368:395-400.
6. Kraemer MUG, Yang CH, Gutierrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020;368:493-7.
7. Deng G, Yin M, Chen X, Zeng F. Clinical determinants for fatality of 44,672 patients with COVID-19. Crit Care 2020;24:179.
8. Docherty AB, Harrison EM, Green CA, et al. Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol. medRxiv 2020:2020.04.23.20076042.
9. Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020;382:1708-20.
10. Khunti K, Singh AK, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of covid-19? BMJ 2020;369:m1548.
11. Parohan M, Yaghoubi S, Seraj A, Javanbakht MH, Sarraf P, Djalali M. Risk factors for mortality of adult inpatients with Coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis of retrospective studies. medRxiv 2020:2020.04.09.20056291.
12. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020;323:2052-9.
13. Wenham C, Smith J, Morgan R, Gender, Group C-W. COVID-19: the gendered impacts of the outbreak. Lancet 2020;395:846-8.
14. Williamson E, Walker AJ, Bhaskaran KJ, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. medRxiv 2020:2020.05.06.20092999.
15. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020;323:1239-42.
16. Yancy CW. COVID-19 and African Americans. JAMA 2020;323:1891-2.
17. Johns Hopkins University of Medicine Coronavirus Resource Center. Maps & trends. 2020, May 22. (Accessed May 22, 2020, at https://coronavirus.jhu.edu/data#charts.)
18. van Bunnik BAD, Morgan ALK, Bessell P, et al. Segmentation and shielding of the most vulnerable members of the population as elements of an exit strategy from COVID-19 lockdown. medRxiv 2020:2020.05.04.20090597.

19. Emanuel EJ, Persad G, Upshur R, et al. Fair allocation of scarce medical resources in the time of Covid-19. Mass Medical Soc; 2020:2049-55.
20. Schnake-Mahl AS, Carty MG, Sierra G, Ajayi T. Identifying Patients with Increased Risk of Severe Covid-19 Complications: Building an Actionable Rules-Based Model for Care Teams. NEJM Catalyst Innovations in Care Delivery 2020;1.
21. Banerjee A, Pasea L, Harris S, et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. The Lancet 2020. DOI: https://doi.org/10.1016/S0140-6736(20)30854-0.
22. Adams ML, Katz DL, Grandpre J. Population based estimates of comorbidities affecting risk for complications from COVID-19 in the US. medRxiv 2020:2020.03.30.20043919.
23. Clark A, Jit M, Warren-Gash C, et al. How many are at increased risk of severe COVID-19 disease? Rapid global, regional and national estimates for 2020. medRxiv 2020:2020.04.18.20064774.
24. Centers for Disease Control and Prevention. Preliminary estimate of excess mortality during the covid-19 outbreak — new york city, march 11–may 2, 2020. 2020, May 15. (Accessed May 22, 2020, at https://www.cdc.gov/mmwr/volumes/69/wr/mm6919e5.htm.)
25. County of Los Angeles. Project Roomkey. 2020. (Accessed May 23, 2020, at https://covid19.lacounty.gov/project-roomkey/.)
26. Nadja Popovich, Anjali Singhvi and Matthew Conlen. Where chronic health conditions and coronavirus could collide. The New York Times, 2020. (Accessed May 22, 2020, at https://www.nytimes.com/interactive/2020/05/18/us/coronavirus-underlying-conditions.html?auth=link-dismiss-google1tap.)
27. Chow N, Fleming-Dutra K, Gierke R, et al. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 — United States, February 12–March 28, 2020. MMWR Morb Mortal Wkly Rep 69(13):382-386 2020;69.
28. NYC Open Data. NYC Open Data. 2020 population. 2020. (Accessed May 24, 2020, at https://data.cityofnewyork.us/City-Government/2020-population/t8c6-3i7b.)
29. Centers for Disease Control and Prevention. Covid-19 forecasts. 2020, May 21. (Accessed May 22, 2020, at https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html.)
30. Campling J, Jones D, Chalmers JD, et al. The impact of certain underlying comorbidities on the risk of developing hospitalised pneumonia in England. Pneumonia (Nathan) 2019;11:4.
31. Doll S, Paccaud F, Bovet P, Burnier M, Wietlisbach V. Body mass index, abdominal adiposity and blood pressure: consistency of their association across developing and developed countries. Int J Obes Relat Metab Disord 2002;26:48-57.
32. Menotti A, Blackburn H, Kromhout D, Nissinen A, Adachi H, Lanti M. Cardiovascular risk factors as determinants of 25-year all-cause mortality in the seven countries study. Eur J Epidemiol 2001;17:337-46.
33. Shea KM, Edelsberg J, Weycker D, Farkouh RA, Strutton DR, Pelton SI. Rates of pneumococcal disease in adults with chronic medical conditions. Open Forum Infect Dis 2014;1:ofu024.
34. Athanasiou AE, Mazaheri M, Zarrinnia K. Dental arch dimensions in patients with unilateral cleft lip and palate. Cleft Palate J 1988;25:139-45.
35. NYC Health. COVID-19: Data. 2020, May 22. (Accessed May 22, 2020, at https://www1.nyc.gov/site/doh/covid/covid-19-data.page.)

36. Centers for Disease Control and Prevention. Covid-19 in racial and ethnic minority groups. 2020. (Accessed May 10, 2020, at https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html.)
37. NYC Health. Age-adjusted rates of lab confirmed COVID-19. 2020. (Accessed May 18, 2020, at https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-race-ethnicity-04162020-1.pdf.)
38. Zaninotto P, Batty GD, Stenholm S, et al. Socioeconomic Inequalities in Disability-free Life Expectancy in Older People from England and the United States: A Cross-national Population-Based Study. J Gerontol A Biol Sci Med Sci 2020;75:906-13.
39. US Census Bureau. American Community Survey (ACS). 2020. (Accessed May 18, 2020, at https://www.census.gov/programs-surveys/acs.)
40. US Census Bureau. ACS 2017 age-sex table. 2018. (Accessed May 9, 2020, at https://data.census.gov/cedsci/table?q=Age%20and%20Sex&hidePreview=true&t=Age%20and%20Sex&tid=ACSST1Y2017.S0101&vintage=2018&y=2017.)
41. US Census Bureau. Hispanic or Latino origin by race. 2018. (Accessed May 9, 2020, at https://data.census.gov/cedsci/table?q=hispanic&hidePreview=true&tid=ACSDT1Y2018.B03002&t=Hispanic%20or%20Latino&vintage=2018.)
42. Centers for Disease Control and Prevention. 500 cities: local data for better health, 2019 release. 2019. (Accessed May 9, 2020, at https://chronicdata.cdc.gov/500-Cities/500-Cities-Local-Data-for-Better-Health-2019-relea/6vp6-wxuq.)
43. Centers for Disease Control and Prevention. United States cancer statistics. 2019. (Accessed May 10, 2020, at https://www.cdc.gov/cancer/uscs/dataviz/download_data.htm.)
44. Robert Graham Center. Social deprivation index (SDI). 2020. (Accessed May 18, 2020, at https://www.graham-center.org/rgc/maps-data-tools/sdi/social-deprivation-index.html.)
45. Centers for Disease Control and Prevention. National health interview survey. 2020. (Accessed May 10, 2020, at https://www.cdc.gov/nchs/nhis/.)

Figure 1: Distribution of the Index of Excess Risk (IER) for COVID-19 mortality across 442 US cities. The index is defined based on risk of mortality for the population at large associated with age, gender, ethnicity, social deprivation index and 12 different health conditions. The index is standardized using a reference value that corresponds to average risk across the cities weighted by their population sizes. Results are shown using histogram (A) and a US geographic map (B). See Methods and Supplementary Notes for the definition of IER.

Figure 2: Distribution of the proportion and number of individuals at different risk-thresholds across 442 US cities. The risk thresholds are defined with respect to average of risks across all the cities weighted by population sizes. The results are obtained using information on prevalence and co-occurrence of the predisposing factors from various national database, and mathematical approximations for the distribution of underlying risk-score using mixture normal distributions. See Methods and Supplementary Notes for details. Results for additional risk-thresholds are shown in Supplementary Figure 6.

Figure 3: Projections for the proportion of deaths expected to occur at various risk-thresholds across 442 US Cities. Histogram of the proportions across the cities (left panel) and scatter plots for them again the corresponding proportion of general populations exceeding the risk-thresholds (right panel) are shown. The projections are based on the underlying mixture normal distribution of the risk-scores in the general population and assuming the association parameters used are valid for the general US population. Results for additional risk thresholds are shown in Supplementary Figure 7.

Table 1. The percentages of NHIS population that exceed various risk-thresholds, overall and by age group. Risk-thresholds are evaluated in reference to the average risk over all subjects. The analysis does not include the measure of social deprivation index (SDI) which is unavailable in NHIS.

	2-fold Risk (Elevated)	3-fold Risk (Substantially Elevated)	5-fold Risk (High)	10-fold Risk (Very High)	25-fold Risk (Extremely High)
Overall	12.31	7.93	4.39	1.43	0.21
Age 18-39	0.00	0.00	0.00	0.00	0.00
Age 40-49	0.06	0.01	0.00	0.00	0.00
Age 50-59	2.09	0.87	0.40	0.00	0.00
Age 60-69	15.83	7.40	2.35	0.27	0.05
Age 70-79	59.70	36.67	16.95	4.02	0.15
Age 80+	100.00	79.57	56.98	23.77	4.43

Figure 1

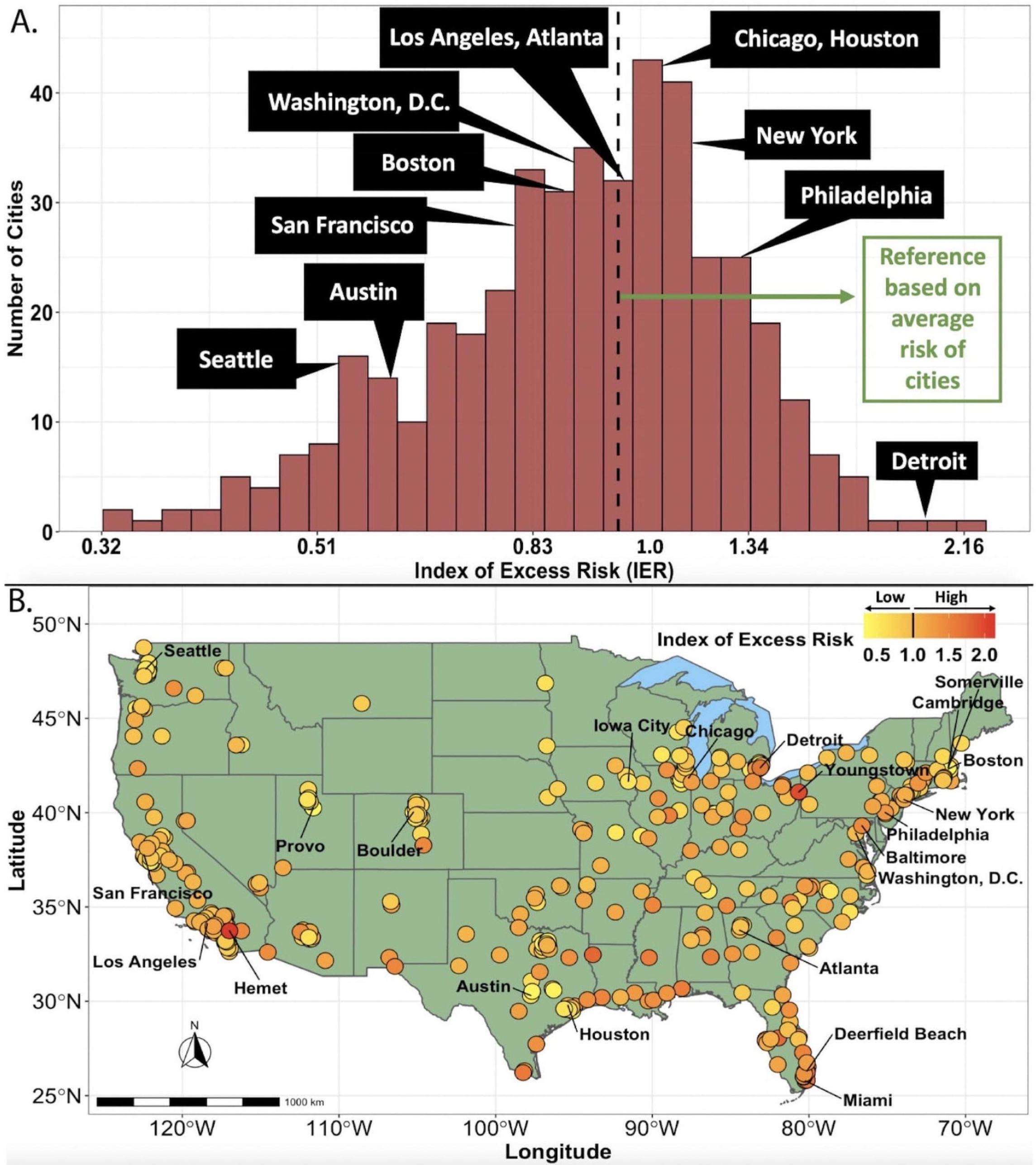

A.
Los Angeles, Atlanta
Chicago, Houston
Washington, D.C.
New York
Boston
Philadelphia
San Francisco
Austin
Seattle
Reference based on average risk of cities
Detroit
Number of Cities
0
10
20
30
40
0.32
0.51
0.83
1.0
1.34
2.16
Index of Excess Risk (IER)
B.
Low
High
Index of Excess Risk
0.5 1.0 1.5 2.0
Seattle
Somerville
Cambridge
Iowa City
Chicago
Detroit
Boston
Youngstown
New York
Philadelphia
Baltimore
Washington, D.C.
Provo
Boulder
San Francisco
Los Angeles
Hemet
Austin
Houston
Atlanta
Deerfield Beach
Miami
N
1000 km
50°N
45°N
40°N
35°N
30°N
25°N
Latitude
120°W
110°W
100°W
90°W
80°W
70°W
Longitude

Figure 2

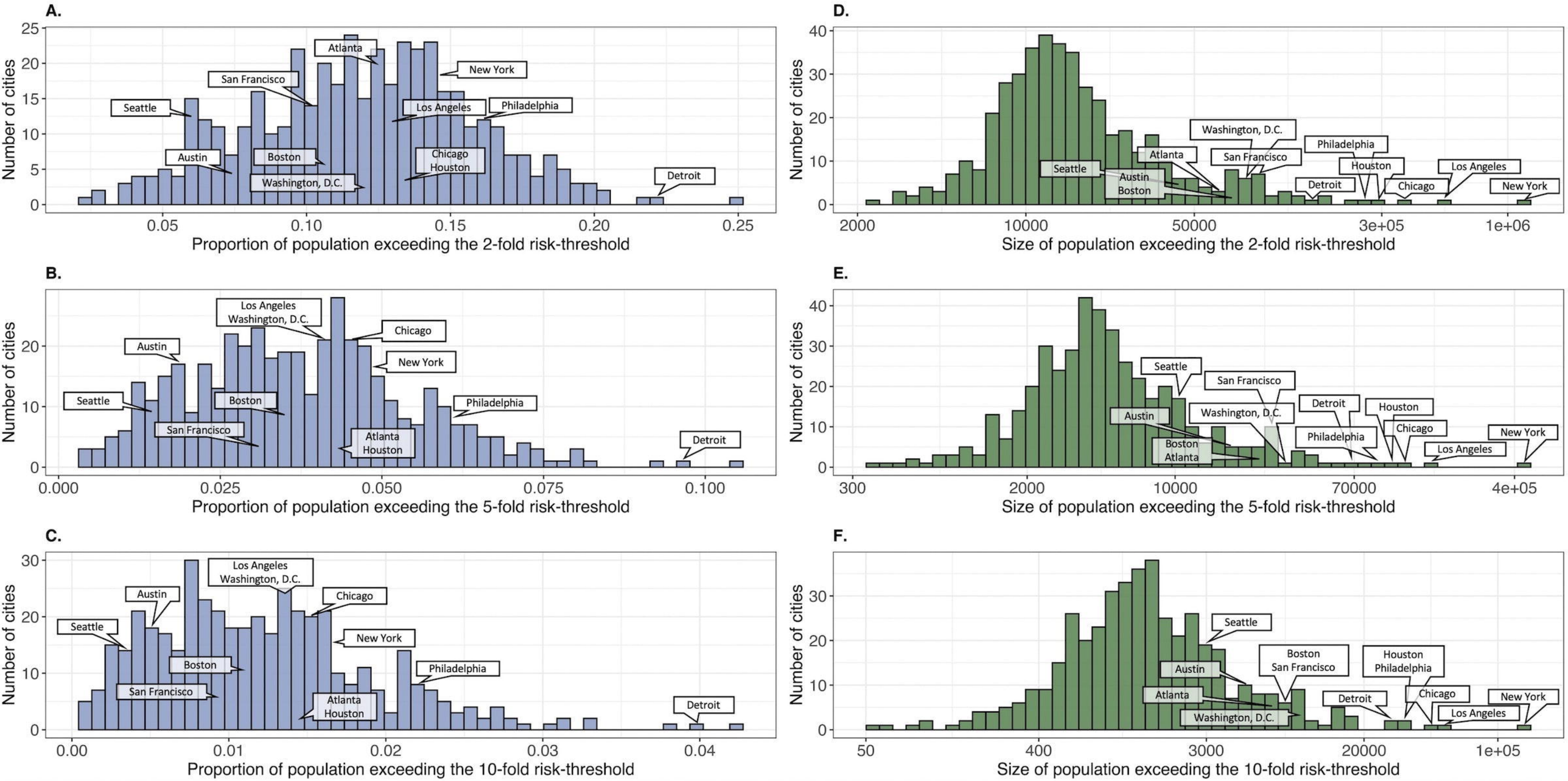
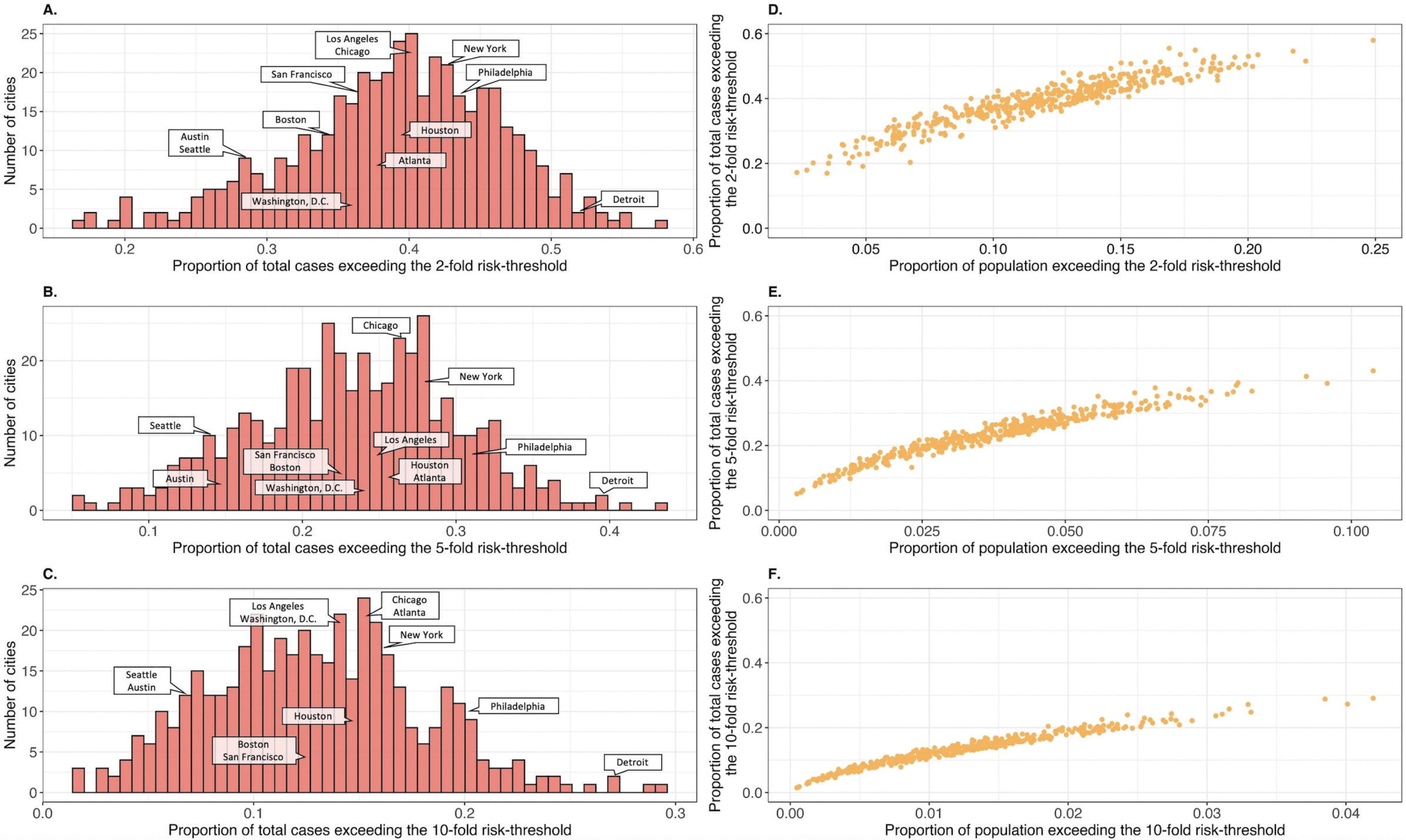

A.
Number of cities
Proportion of population exceeding the 2-fold risk-threshold
Atlanta
New York
San Francisco
Seattle
Los Angeles
Philadelphia
Austin
Boston
Chicago
Houston
Washington, D.C.
Detroit
B.
Number of cities
Proportion of population exceeding the 5-fold risk-threshold
Los Angeles
Washington, D.C.
Chicago
Austin
New York
Seattle
Boston
Philadelphia
San Francisco
Atlanta
Houston
Detroit
C.
Number of cities
Proportion of population exceeding the 10-fold risk-threshold
Los Angeles
Washington, D.C.
Austin
Chicago
Seattle
New York
Boston
Philadelphia
San Francisco
Atlanta
Houston
Detroit
D.
Number of cities
Size of population exceeding the 2-fold risk-threshold
Washington, D.C.
Philadelphia
Atlanta
San Francisco
Seattle
Houston
Los Angeles
Austin
Boston
Detroit
Chicago
New York
E.
Number of cities
Size of population exceeding the 5-fold risk-threshold
Seattle
San Francisco
Detroit
Houston
Austin
Washington, D.C.
Chicago
New York
Philadelphia
Boston
Atlanta
Los Angeles
F.
Number of cities
Size of population exceeding the 10-fold risk-threshold
Seattle
Boston
San Francisco
Houston
Philadelphia
Austin
Atlanta
Detroit
Chicago
New York
Washington, D.C.
Los Angeles

Figure 3


A.
Number of cities
Proportion of total cases exceeding the 2-fold risk-threshold
Los Angeles
Chicago
New York
San Francisco
Philadelphia
Boston
Houston
Austin
Seattle
Atlanta
Washington, D.C.
Detroit
B.
Number of cities
Proportion of total cases exceeding the 5-fold risk-threshold
Chicago
New York
Seattle
Los Angeles
Philadelphia
San Francisco
Boston
Houston
Atlanta
Austin
Washington, D.C.
Detroit
C.
Number of cities
Proportion of total cases exceeding the 10-fold risk-threshold
Los Angeles
Washington, D.C.
Chicago
Atlanta
New York
Seattle
Austin
Philadelphia
Houston
Boston
San Francisco
Detroit
D.
Proportion of total cases exceeding the 2-fold risk-threshold
Proportion of population exceeding the 2-fold risk-threshold
E.
Proportion of total cases exceeding the 5-fold risk-threshold
Proportion of population exceeding the 5-fold risk-threshold
F.
Proportion of total cases exceeding the 10-fold risk-threshold
Proportion of population exceeding the 10-fold risk-threshold