

How do pandemics end? Two decades of recurrent outbreak risk following the main waves

Max Schroeder

University of Glasgow <https://orcid.org/0000-0002-5425-5645>

Spyridon Lazarakis

Lancaster University <https://orcid.org/0000-0001-7044-6110>

Rebecca Mancy (✉ Rebecca.Mancy@glasgow.ac.uk)

University of Glasgow <https://orcid.org/0000-0002-4667-9229>

Konstantinos Angelopoulos

University of Glasgow <https://orcid.org/0000-0002-0702-3673>

Research Article

Keywords: pandemics, outbreak risk, Spanish flu, Russian flu, COVID-19

Posted Date: September 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-893004/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

How do pandemics end? Two decades of recurrent outbreak risk following the main waves

Max Schroeder¹, Spyridon Lazarakis², Rebecca Mancy^{3, 4}, Konstantinos Angelopoulos^{1, 5}

Summary

We analyse the dynamic evolution of disease outbreak risk after the main waves of the 1918-19 “Spanish flu” pandemic in the US and in major cities in the UK, and after the 1890-91 “Russian flu” pandemic in England and Wales. We compile municipal public health records and use national data to model the stochastic process of mortality rates after the main pandemic waves as a sequence of bounded Pareto distributions with an exponentially decaying tail parameter. In all cases, we find elevated mortality risk lasting nearly two decades. An application to COVID-19 under model uncertainty shows that in 90% of model-predicted time series, the annual probability of outbreaks exceeding 500 deaths per million is above 20% for a decade, remaining above 10% for two decades.

¹ Adam Smith Business School, University of Glasgow, Glasgow, UK, ² Management School, Lancaster University, Lancaster, UK, ³ MRC/CSO Social & Public Health Sciences Unit, University of Glasgow, Glasgow, UK, ⁴ Institute of Biodiversity, Animal Health and Comparative Medicine, Glasgow, UK ⁵ CESifo, Munich, Germany

 email: Rebecca.Mancy@glasgow.ac.uk

Introduction

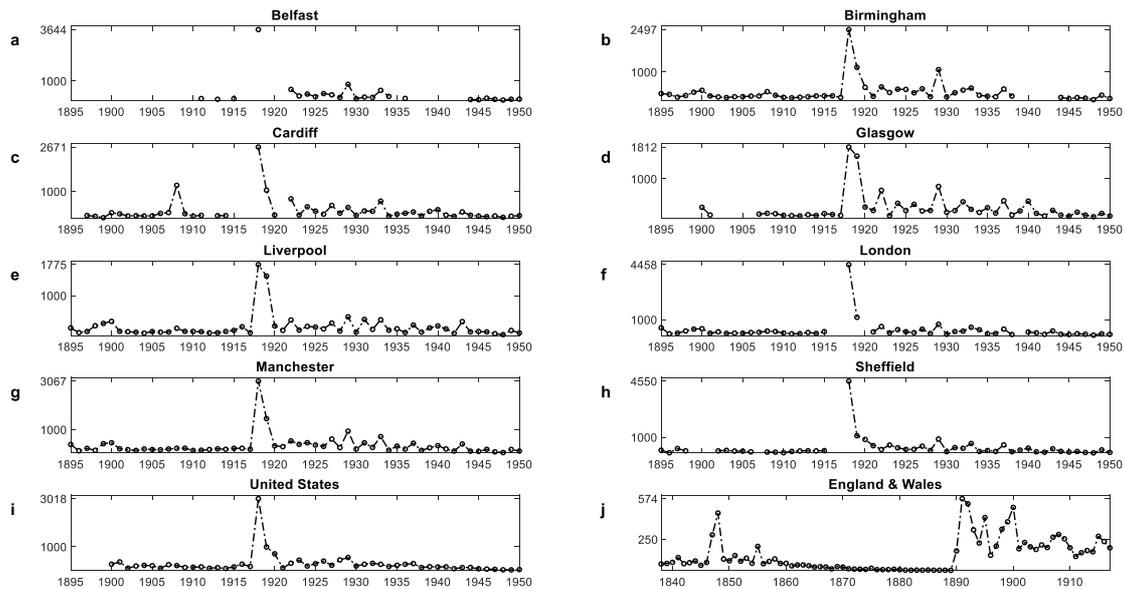
Pandemics are large negative system-wide shocks, affecting humans via increases in health and income risk, as well as reductions in income and welfare.¹⁻⁵ However, pandemics are not one-off events. They typically have several waves spanning more than one year. Moreover, following the main waves, recurrent outbreaks may occur due to re-introduction of the virus, new variants, waning immunity, human behaviour (e.g. vaccine refusal), or population turnover leading to reductions in population-level immunity.⁶⁻⁹ Researchers from a range of relevant disciplines have expressed concern about increased mortality and disease outbreak risk post-COVID-19 in academic outlets^{6,7,10} and in the media.^{11,12}

The potential for elevated mortality, and sizeable disease outbreaks, after the main waves of the pandemic, combined with their importance for health and socioeconomic outcomes,^{1,2,13-15} imply that empirical evidence on medium-run mortality risk is required to aid the modelling of its health and economic implications. A characterisation of mortality risk dynamics is essential to understand the risk exposure that societies and economies face after exiting the main pandemic waves and the speed at which this risk fades in the decades that follow the pandemic. Evidence from pandemics from a period without the recent advances in medicine, public health and technology is particularly valuable in this respect. It is important because it can shed light on the underlying epidemiological risk after the main waves of the pandemic before the effect of modern means of intervention, and thus highlight the potential gains from such preventive and mitigating intervention.

Post-pandemic mortality rates

To analyse mortality risk in the period following the main waves of previous pandemics, we compiled data on mortality rates from influenza for eight large municipalities from across the UK, for England and Wales combined, and for the US (Figure 1). We collected city-level data on annual influenza mortality rates between 1895 and 1950 from public health records kept at the municipal level in the UK, namely the Medical Officer for Health (MOH) reports (see Supplementary Material A for more details).

Figure 1: Annual mortality rate from influenza



Mortality rate in deaths per million population. The upper y-axis tick indicates the maximum mortality rate for each data series. Data for UK cities are taken from Medical Officer for Health reports. Data for US is taken from Lindner and Grove (1943)¹⁶ and Grove and Hetzel (1968).¹⁷ Data for England & Wales are taken from Langford (2002),¹⁸ Table 5. Information on population sizes and densities is provided in Table S1.

Focusing first on the eight large British municipalities, the main insight from Figure 1 is that mortality rates are higher and more variable for a long period after the 1918-19 pandemic than prior to it. Although the most striking feature is the massive increase in mortality in 1918 and 1919 due to the main pandemic waves, we also observe spikes of high mortality in the series post-1920 that continue for an extended period. This feature implies increased disease outbreak risk. Although rates during these outbreak years are lower than those of 1918 and 1919, they are significantly higher than those before 1918 and those two decades later. Contemporary expert evaluations by the Medical Officer for London also identify 1922, 1924, 1927, 1929, 1933 and 1937 as years of elevated mortality from influenza (see Supplementary Material A).

We also examine post-pandemic disease outbreaks using national-level data. Figure 1 also shows the mortality rate from influenza for the US between 1900 and 1950. The same pattern as that observed for the UK municipalities emerges. Finally, using data on influenza mortality for England and Wales, we find a similar pattern of increased disease outbreaks following the aftermath of the 1890-91 "Russian Flu" pandemic. This pandemic began in the latter part of 1889 and claimed over 1 million lives worldwide, making it one of the most significant global pandemics of the time.^{19,20} Revisiting the first four rows of Figure 1, the spikes in mortality in the last decade of the 19th century and first decade of the 20th century are thus probably related to the increased risk following the 1890-91 pandemic.

Dynamics of mortality risk after the main pandemic waves

Here we conduct modelling analysis of the dynamic evolution of mortality risk after the main pandemic waves, particularly the level of disease outbreak risk and the speed at which it fades, conditional on the mortality effect of the main

waves. We model the time series of mortality rates after the main waves of a pandemic as outcomes drawn from a sequence of bounded Pareto distributions such that mortality risk and the probability of large outbreaks declines over time relative to the impact of the initial waves. We choose the bounded Pareto distribution to model mortality risk in any given year because the size of epidemics and outbreaks has been shown to be highly over-dispersed, and is therefore well modelled by a fat-tailed distribution such as a bounded Pareto.²¹ We focus on the period following the main pandemic waves and assume that the inverse of the tail index of the bounded Pareto distributions decays exponentially after the main waves of the pandemic. The mortality effect of the main waves is reflected in the choice of the bounds of the Pareto distribution. Denote the mortality rate in year t by d_t , for $t = 0, 1, 2, \dots, N$, where the time period refers to (1920, 1921, ..., 1950) for the eight cities in the UK and the US with reference to the 1918-19 pandemic, and to (1892, 1893, ..., 1917) for England and Wales with reference to the 1890-91 pandemic. In each year, mortality rates are drawn from a bounded Pareto distribution

$$d_t \sim PD_t(d_l, d_u, \alpha_t), \quad (1)$$

where $\alpha_t > 0$ is the time varying Pareto tail index, and $d_l > 0$ and $d_u > d_l$ are, respectively, lower and upper bounds. Defining $\eta_t = 1/\alpha_t$, we assume that

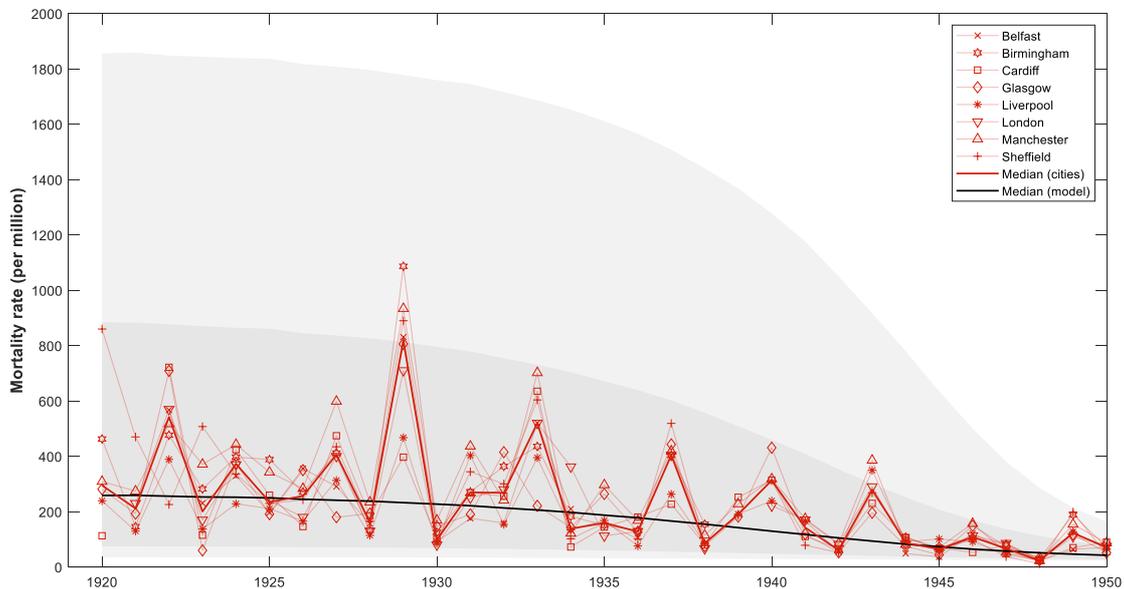
$$\eta_t = \eta_0 e^{-\lambda t}, \quad (2)$$

where $\lambda > 0$ determines the rate at which the inverse tail index decays over time, while η_0 sets the initial level of the probabilities by providing the initial value. Conditional on the time process in (2), and thus conditional on the sequence $(\alpha_t)_{t=0}^N$, d_t is independently distributed over time following (1). We fit the model to the data after the main pandemic waves, taking as given the main pandemic event and its experience. Compared with the main pandemic event, recurrent outbreaks imply lower mortality, so that we set the upper bound of mortality rates to the mortality rate of 1918 in each city. Similarly, the lower bound of mortality is determined by the lowest mortality experienced in the longer run in each city, i.e. background influenza mortality. As explained in Supplementary Material B, our main results are robust to modelling d_l and d_u as the theoretical upper and lower bounds of mortality rates, but by exploiting information pertinent to the pandemic in parameterising d_l and d_u the model provides more accurate mortality risk predictions. Moreover, by letting these bounds differ between cities, we allow for differences in the mortality of the main pandemic waves, as well as in latent socioeconomic and public health conditions, to influence risk dynamics. Therefore, the modelling allows for heterogeneity in risk dynamics between geographical units and periods. In turn, this permits applications to predict medium-run mortality risk after the main waves of other pandemics, by conditioning on an upper and lower bound that are relevant for the pandemic and disease in question. Models are fitted to the data by choosing the parameters to maximise the likelihood function (see Supplementary Material B).

Figure 2 shows the time series of data points overlaid on model simulations, confirming that most observations are contained within the interquartile range of simulated outcomes. Observed outcomes outside these bounds are indeed

those of the rarer events: the very large outbreak of 1929, and outbreaks in the later decades that are thus less affected by the modelled pandemic. Model predictions for the US, and England and Wales following the 1890-91 pandemic are also consistent with this pattern (Supplementary Material C, Figure S10).

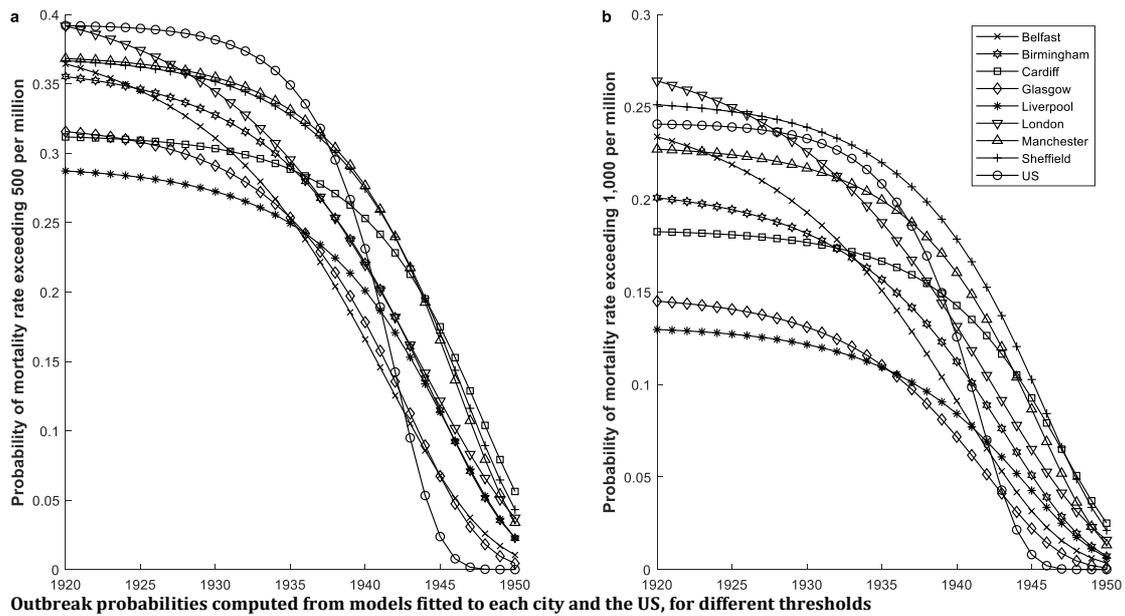
Figure 2: Model predicted influenza mortality rates following the 1918-19 pandemic



Simulated median (solid black line), interquartile range (dark shading) and 80% prediction interval (light shading) are based on 1m random draws. Simulated outcomes are based on the model fitted using the average mortality rates across cities. Data are overplotted in red. Data for UK cities are taken from Medical Officer for Health reports.

Fitting the models to the data, we obtain the time series of distributions of mortality rates between 1920 and 1950. From these, we can calculate the probability of specific mortality rates in a given year. In particular, we can compute the time series of disease outbreak risk by defining a disease outbreak as mortality above a threshold. For the 1918-19 pandemic, we define two disease outbreak thresholds at 500 or at 1000 deaths per million. The first is about one-third the rate of the main waves (1918 and 1919) and significantly higher than any mortality rate between 1900 and 1918 (i.e. after the 1890-91 pandemic effects had died out). A death rate of 500 per million corresponds to that observed during the main waves of the 1890-91 pandemic and its recurrent outbreaks. This threshold identifies as outbreaks in the data post-1920, the same years as those described as being of exceptionally high mortality in the MOH reports for London. The 1000 threshold corresponds to recurrent disease outbreaks approaching the levels experienced in some cities during the main pandemic waves and approximately that of the severe 1929 outbreak. More generally, the 1000 threshold identifies particularly severe disease outbreaks that were *possible* given the dynamic process for mortality risk, even if unrealised *ex post*. Similar patterns are observed for the 1890-91 pandemic in England and Wales, which had lower mortality (Supplementary Material C, Figure S11).

Figure 3: Outbreak risk following the 1918-19 pandemic



We plot the post-1918-19 disease outbreak probabilities in Figure 3. Two main qualitative results stand out. First, the probability of a disease outbreak remains high for about two decades after the main pandemic waves. Disease outbreak risk does not decline exponentially, but instead there is a very low rate of decline for a prolonged period. During the transition to background influenza mortality, outbreaks with severe mortality can occur with relatively high probability. Second, the pattern of the time evolution of risk is similar across all cities and for the US. This pattern holds despite considerable differences between the geographical units that we study in the severity of the impact of the pandemic during the main waves (Figure 1) and in differences in background influenza mortality. Comparing the results for the two different thresholds, the dynamic pattern for all geographical units and the two pandemics in our data remains, despite the naturally higher outbreak probabilities for the lower threshold.

Discussion: lessons from the past for the present

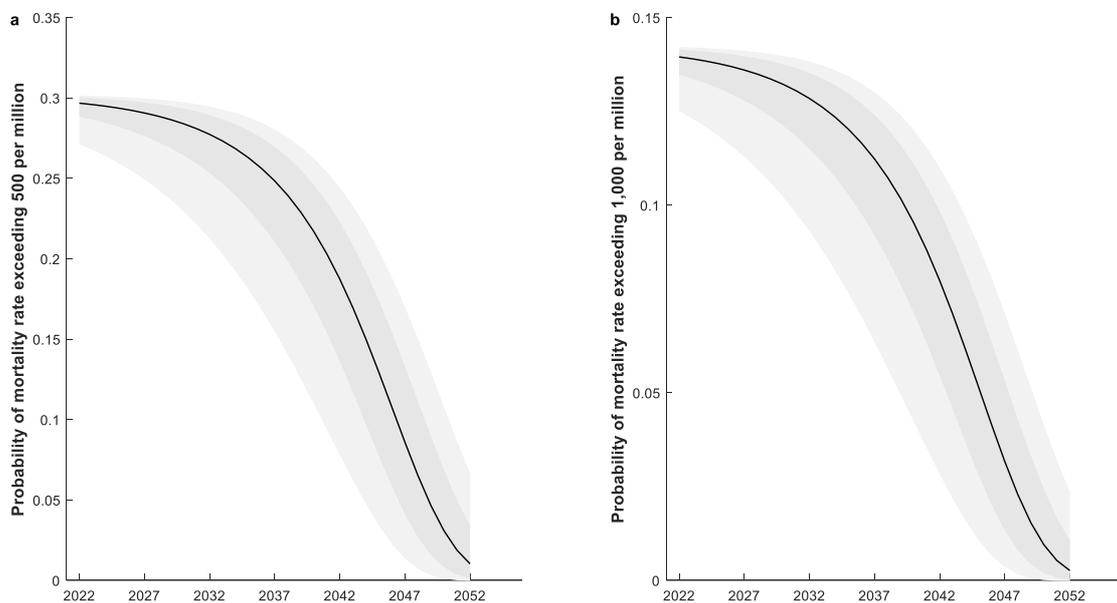
The similarity of the pattern of disease outbreak dynamics across geographic units with different socioeconomic characteristics and two historical pandemics highlights the generality of our finding that mortality risk remains elevated for a prolonged period after the main pandemic waves and motivates using the statistical model to predict mortality after the main waves of pandemics in the 21st century. The world in the 21st century is, if anything, more connected and densely populated than the world of more than one hundred years ago. Therefore, it is likely that without systematic intervention and the benefit of medical and scientific progress, pandemics in the modern world also have the potential to lead to a period of high and persistent mortality and disease outbreak risk.

We use the historical mortality risk estimates to simulate mortality dynamics after the main waves of COVID-19, allowing for model uncertainty. Figure 4 shows model predictions regarding the probabilities of disease outbreaks (mortality rates exceeding 500 and 1000 per million) from a counterfactual

analysis of disease outbreak risk after the main waves of COVID-19. We set the upper bound on future mortality to be determined by mortality in the UK in 2020 and the lower bound to reflect mortality in the non-pandemic state in the coming decades, which we approximate by background influenza mortality pre-2020. The results demonstrate elevated disease outbreak risk for two decades.

To obtain the model predictions in Figure 4, we allow for model uncertainty. In particular, we allow for uncertainty about which sequence of distributions in $(PD_t)_{t=0}^N$ generates the data; this uncertainty is different from the epidemiological uncertainty regarding mortality rates that is generated by a given sequence of distributions. To account for uncertainty regarding the dynamic path of disease outbreak risk (given the bounds for the Pareto distributions), we use the range of values of η_0 and λ across the different geographical units to approximate the joint distribution of possible parameter values (see Supplementary Material B). Then, we calculate the disease outbreak probabilities for one million draws and plot in Figure 4 relevant percentiles of the distribution of disease outbreak probabilities over possible data generating processes. In 90% of model-predicted dynamic paths, the probability of outbreaks exceeding 500 deaths per million, which is 20 times higher than that of seasonal influenza, is above 20% for a decade and remains above 10% for two decades. Regarding outbreaks that come closer to the size of the main wave mortality, we find that in 90% of model-predicted dynamic paths, the probability of outbreaks exceeding 1000 deaths per million is above 10% for a decade.

Figure 4: Predicted outbreak risk following the 2020-21 pandemic



Simulated median (solid black line), interquartile range (dark shading) and 80% prediction interval (light shading) are based on 1m random draws. Outbreak probabilities computed from model parameterisations that are drawn from a distribution of η_0 and λ implied by the models fitted to historical data ($d_u = 1858$, $d_l = 24$).

Our findings can inform research to understand the implications of pandemics for a range of health and socioeconomic outcomes by quantifying post-pandemic mortality risk and its persistence over time. Disease outbreak risk generates health and economic uncertainty which has significant and unequal

consequences for socioeconomic outcomes across the population. Given that disease outbreaks imply a deterioration in health outcomes that is not symmetric across the population^{1,2,13,14}, a period of increased disease outbreak risk implies the possibility of repeated negative shocks to health and health inequality. Moreover, outbreak risk generates social, political and economic uncertainty, stemming from possible impacts of the disease itself and of containment measures on economic activity. Increased aggregate-level uncertainty impacts economic decision-making, income inequality and economic fluctuations, either directly or via increased idiosyncratic risk.²²⁻²⁶

Medicine, public health, technology, and better-informed and prepared policy intervention offer the opportunity in the modern world to confront post-pandemic recurrent outbreak risk better than a century ago. In this sense, our findings regarding the potential severity of disease outbreak risk for many years after a pandemic highlights the value of scientific and medical developments to mitigate these, and more generally of prevention and policy preparedness to mitigate disease outbreaks even after exiting the main pandemic waves.

Supplementary Material A

Data

City-level data were compiled from Medical Officer of Health (MOH) reports. These reports were annual administrative documents covering a range of public health-related issues at the municipal level. The first reports begin in the mid-19th century, and coverage extends to most municipalities in the UK until the early 1970s. The reports used here have been digitised and can be viewed on the Wellcome Trust Collection website.²⁷

We collected influenza mortality data between 1895 and for eight large municipalities from across the UK: Belfast, Birmingham, Cardiff, Glasgow, London, Liverpool, Manchester and Sheffield. To do this, we searched relevant MOH reports for each municipality. Generally, annual mortality rates by cause of death are presented in tables within the report or its appendices (or could be computed as the ratio of the number of deaths to the population size). Despite changes to the taxonomy of many causes of death over time, *influenza* was reported in the MOH reports throughout the decades. During the war years, 1914-1919 and 1939-1945, some reports are missing, or the data provided is incomplete. In these cases, we recovered the missing information by assessing statistics for these years from later reports, where possible. The dataset collected from the MOH reports underlying this research paper is available on GitHub at <https://github.com/maxschr90/Schroeder-et-al.-2021--How-long-do-pandemics-last->.

The narrative provided in the MOH reports for London County Council confirms that the two decades after the three initial waves of the pandemic were characterised by several further large influenza outbreaks. In particular, we examined available London County Council MOH reports between 1920 and 1957 to identify discussions of increased prevalence of influenza in a given year. In each report, we performed a search for the phrase *influenza*. As public health officials were particularly attentive to influenza after the 1918-19 pandemic, virtually every annual report contains at least some discussion of the disease. Generally, the language of the reports is clear on whether a certain year is considered to have a notable outbreak of influenza. Some reports further included retrospective reflections on past influenza outbreaks. The reports identify 1922, 1924, 1927, 1929, 1933 and 1937 as years of heightened mortality from influenza relative to other years.

Summary statistics on population and population density for the UK cities are in Table S1.

Table S1: Population and population density for UK cities

	Population			Population density		
	1920s	1930s	1940s	1920s	1930s	1940s
Belfast	422,130	425,202	442,935	-	-	-
Birmingham	947,923	1,023,811	1,053,157	21	20	21
Cardiff	219,894	222,658	227,861	21	18	17
Glasgow	1,079,858	1,088,829	1,089,368	51	35	27
Liverpool	837,595	858,783	720,112	40	32	-
London	4,540,000	4,230,295	2,863,548	-	-	-
Manchester	749,970	746,974	641,040	35	29	24
Sheffield	519,224	517,967	492,093	17	14	13

Population and population density numbers refer to decadal averages compiled from available data in the MOH reports. Population density is measured in persons/acre.

Data for the US are taken from the annual vital statistics reports compiled by the [National Center for Health Statistics](#). Specifically, we rely on the two special volumes^{16,17} covering the period 1900-1960. Both volumes are available on the CDC's website.²⁸ Data for England and Wales between 1838 and 1917 are taken from Langford (2002),¹⁸ Table 5, who compiles mortality rates from different sources.

To obtain COVID-19 mortality for the analysis in Figure 4, we combined the total number of UK COVID-19 deaths between 6th March 2020 and 6th of March 2021 (124,654).²⁹ With mid-year population data from the ONS for 2020 (67,081,000).³⁰ This implies a mortality rate of 1,858 per million. Pre-2020 deaths from influenza are taken from the 2018-19 total mortality figures for England & Wales from the ONS³¹ together with the mid-2018 population figures for England & Wales.³²

Supplementary Material B

Modelling of mortality risk dynamics

The model in (1) – (2) assumes that mortality rates after the main waves of the pandemic are the outcomes of a sequence of bounded Pareto distributions, where the inverse of the tail index of these distributions decays exponentially over time. The parameters $d_l > 0$ and $d_u > d_l$ scale the range of mortality rates that the model predicts. We fit the model to the data for each geographical unit and conditional on its experience of the pandemic. Hence, we choose the bounds to reflect the realised range of mortality rates for the geographical unit over the period modelled. Conditional on d_l and d_u , the two parameters η_0 and λ then determine the dynamics of mortality and disease outbreak risk via controlling the level and time evolution of probabilities of outcomes associated with the tail of the Pareto distributions.

We obtain λ and η_0 by maximising the likelihood function:

$$L(\lambda, \eta_0) = \prod_{t=0}^N \frac{(\eta_0 e^{-\lambda t})^{-1} (d_l)^{(\eta_0 e^{-\lambda t})^{-1}} (\bar{d}_t)^{-(\eta_0 e^{-\lambda t})^{-1} - 1}}{1 - \left(\frac{d_l}{d_u}\right)^{(\eta_0 e^{-\lambda t})^{-1}}}, \quad (3)$$

given a sample of mortality rates $(\bar{d}_t)_{t=0}^N$. To maximise the log likelihood function, we use MATLAB's `fmincon` routine, using a sequential quadratic programming algorithm. Derivatives are approximated by central numerical derivatives, and the relevant termination criteria are set to $1e^{-12}$. To account for potential nonconvexities and the presence of local maxima, we begin the maximisation from 1000 random seed values. The estimated parameters are summarised in Table S2.

Table S2: Base model parameters

	λ	η_0	d_l	d_u
City average	0.151	76.867	24	3059
Belfast	0.119	24.859	22	3644
Birmingham	0.144	58.442	30	2497
Cardiff	0.184	360.214	13	2671
Glasgow	0.166	74.119	33	1812
Liverpool	0.166	126.778	23	1775
London	0.11	26.352	23	4458
Manchester	0.18	208.861	23	3067
Sheffield	0.169	202.69	12	4550
US	0.278	802.709	31	3018
England & Wales	0.111	7.684	113	574

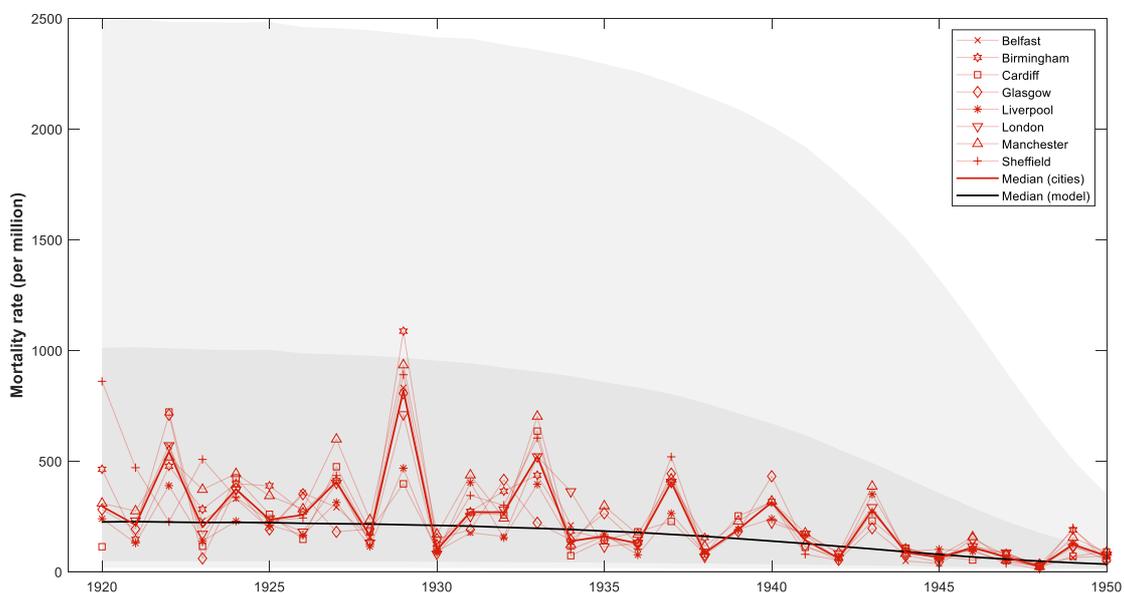
The model in (1) – (2) allows mortality risk to vary with time step t after the main pandemic waves, as it declines towards background mortality. The modelling is motivated by the empirical observations discussed under Figure 1: the period after the pandemic has higher variation in mortality, and thus a higher mortality risk, than background mortality (i.e. during the period before the pandemic and the period a few decades after the pandemic). These observations imply that mortality risk just after the main pandemic waves should start from a higher level, converging to background mortality. Under the assumption that risk is decreasing over time, the model in (1) – (2) describes the level and rate of decline as a function of the parameters η_0 and λ , which determine the sequence $(\alpha_t)_{t=0}^N$. Although the model constrains the inverse of the tail index to decay exponentially, it does not impose such a restriction on *probabilities* associated with the tails. As determined by (1), the probabilities of different mortality rates are a monotonic function of the tail index and thus decrease over time. However, the inverse of a convex function does not have a predetermined convexity. Hence, $\lambda > 0$ implies that tail probabilities decline over time, but not that their rate of decline is exponential. This feature is evident in the results.

The choice of the bounds, d_l and d_u , scales the range of mortality rates. In the baseline application, d_u for each city is set to the mortality rate of 1918 for that city, and d_l to lowest mortality experienced in the long run in each city, or the

average across cities for the model fitted to the average. For the US following the 1918-19 pandemic, d_l and d_u are the lowest post-1918 and the 1918 mortality rates observed in the US time series; for the 1890-91 post-pandemic mortality rate dynamics, these parameters refer to the lowest between 1891 and 1917, and the 1891 mortality rates in the respective time series for England and Wales. These parameters are also summarised in Table S2.

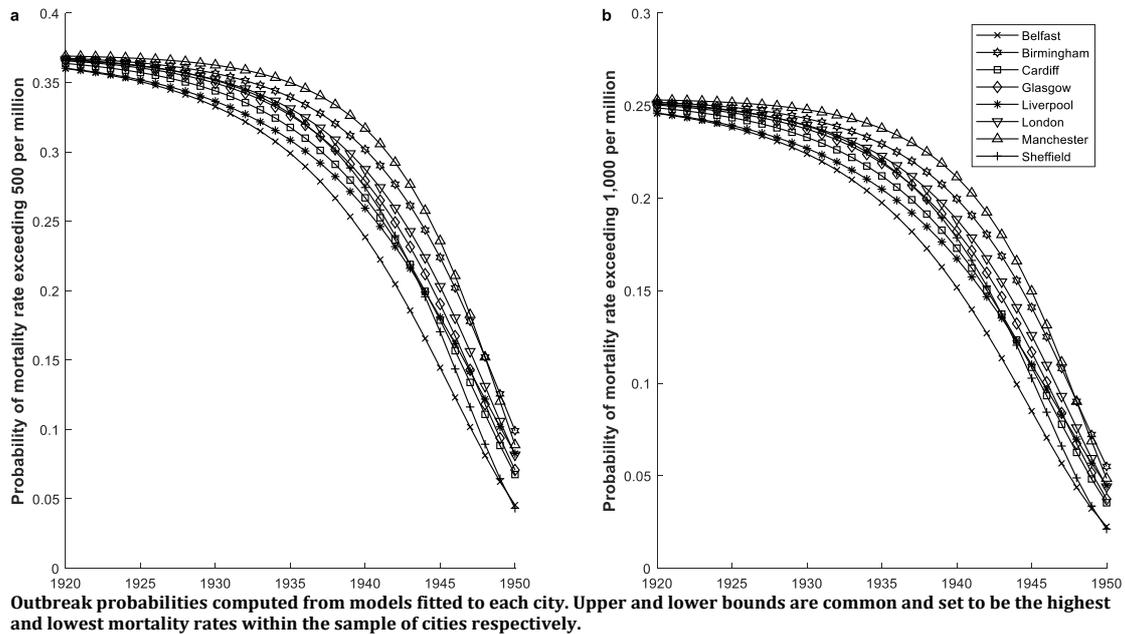
The advantage of letting d_l and d_u be specific to the geographical unit studied is that we allow for differences in the experience of the pandemic and in conditions that influence background infectious disease mortality to have a bearing on risk dynamics. By allowing for such latent variation (see, e.g. Table S1 in Supplementary Material A for variation in population and density across the cities in the UK), our finding of a general pattern of mortality and disease outbreak risk dynamics is stronger. In Figure S1-S2, we also present results for a model of the different cities in which we set d_l and d_u to be common across all cities, specifically to the lowest and highest mortality rate observed across them. As can be seen, the overall pattern is very similar across cities (compared with the results in Figures 2 and 3, even greater similarity is observed). However, the finding of generality as based on Figure S2 on its own is likely biased because of omitted variation between the cities, which can lead to excess similarity. In contrast, the results in Figures 2 and 3 reassure us that the general pattern remains similar even after controlling for variation between cities.

Figure S1: Model predicted influenza mortality rates following the 1918-19 pandemic (model with common bounds)



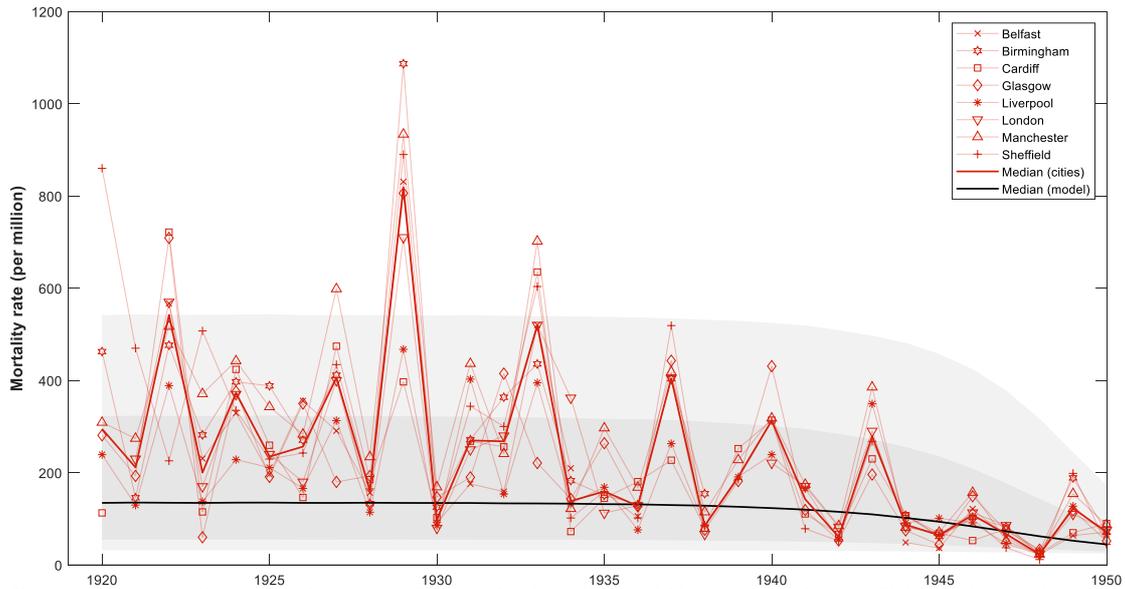
Simulated median (solid black line), interquartile range (dark shading) and 80% prediction interval (light shading) are based on 1m random draws. Simulated outcomes are based on the model fitted using the average across cities. Upper and lower bounds are common and set to be the highest and lowest mortality rates within the sample of cities respectively. Data are overplotted in red. Data for UK cities are taken from Medical Officer for Health reports.

Figure S2: Outbreak risk following the 1918-19 pandemic (model with common bounds)



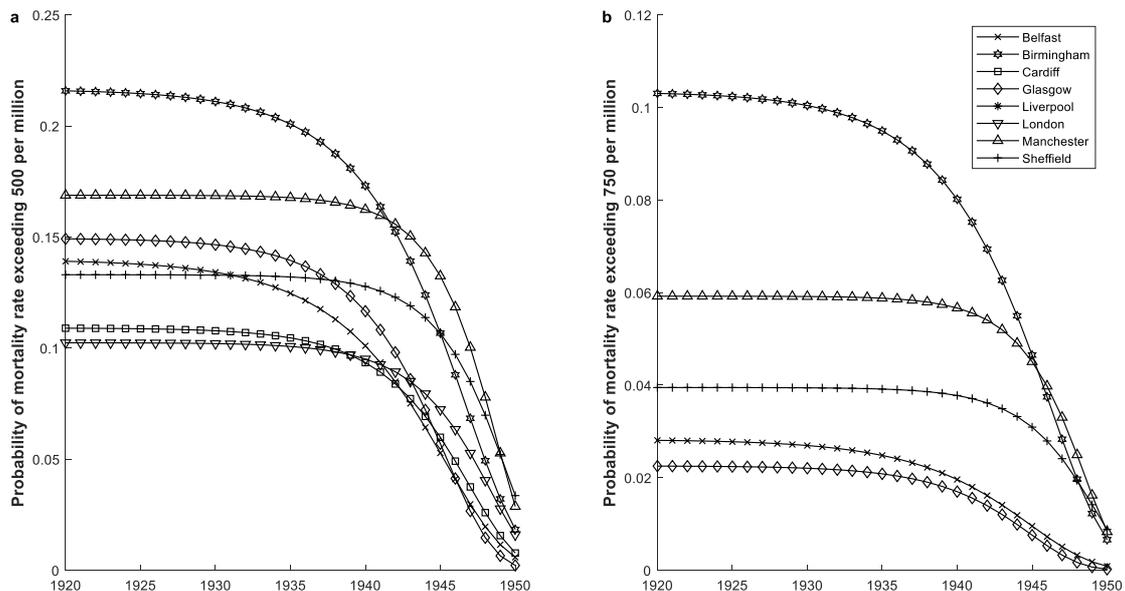
Technically, d_l and d_u can be estimated jointly with η_0 and λ using the time series following the main pandemic waves (i.e., without exploiting the information on the mortality effect of the main pandemic waves). In this case, (3) is maximised by choosing all four parameters, following the same optimisation methods as for the model in (1) – (2). Figures S3 and S4 Figures 2 and 3 under this approach. As can be seen, the model predicted mortality is close to the actual path of mortality, while disease outbreak probabilities for large outbreaks are effectively zero. Although fitting the model in (1) – (2) by estimating all four parameters from the time series of data following the main pandemic waves may be useful for some applications (e.g. if the interest is in summarising historical experience), conceptually, this is no longer a model of mortality and disease-outbreak *risk* dynamics, i.e. of the time evolution of the probability of disease outbreaks. In this case, d_l and d_u are interpreted as parameters chosen to maximise the fit of the process to the data, and thus are chosen by the optimisation routine to be in effect the maximum and minimum mortality observed between 1920 and 1950. In turn, this rules out the possibility of disease outbreaks that are higher than those observed *ex post*, even if theoretically they could have happened. An analysis of risk dynamics must account for the possibility of higher mortality than that that was actually observed. Our modelling approach views bounds as *possible* even if unrealised mortality rates, given the experience of the pandemic, which determines the upper bound, and given the expectation about steady state mortality that is unaffected by the pandemic, which determines the lower bound.

Figure S3: Model predicted influenza mortality rates following the 1918-19 pandemic (model with estimated bounds)



Simulated median (solid black line), interquartile range (dark shading) and 80% prediction interval (light shading) are based on 1m random draws. Simulated outcomes are based on the model fitted using the average across cities. Upper and lower bounds are fitted to the data. Data are overplotted in red. Data for UK cities are taken from Medical Officer for Health reports.

Figure S4: Outbreak risk following the 1918-19 pandemic (model with estimated bounds)

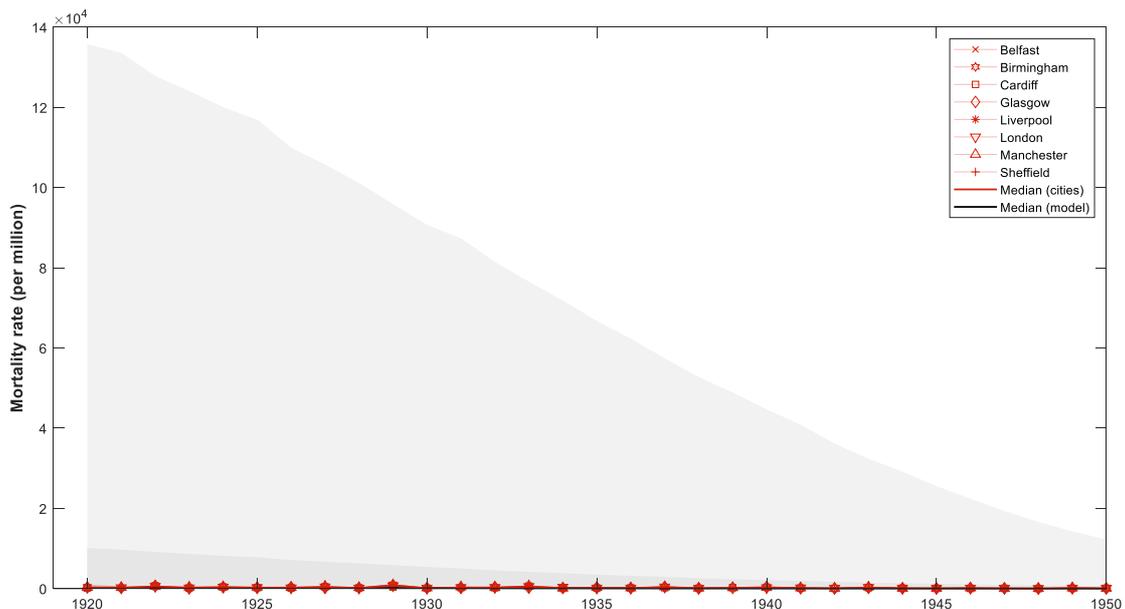


Outbreak probabilities computed from models fitted to each city. Upper and lower bounds are fitted to the data (1920-1950) for each city.

We next compare the approach of specifying the bounds of the Pareto distribution conditionally on the mortality range relevant to the geographical unit to a more agnostic approach of setting d_u and d_l to be determined by theoretical upper and lower bounds of mortality rates. Figures S5 and S6 in reproduce Figures 2 and 3 under the agnostic modelling approach. As can be seen, the pattern in Figure S5 is qualitatively similar with that in Figure 2, but the model predicted mortality rates can be unrealistically high with significant probabilities. Moreover, while in Figure S6 we see that the probability of a

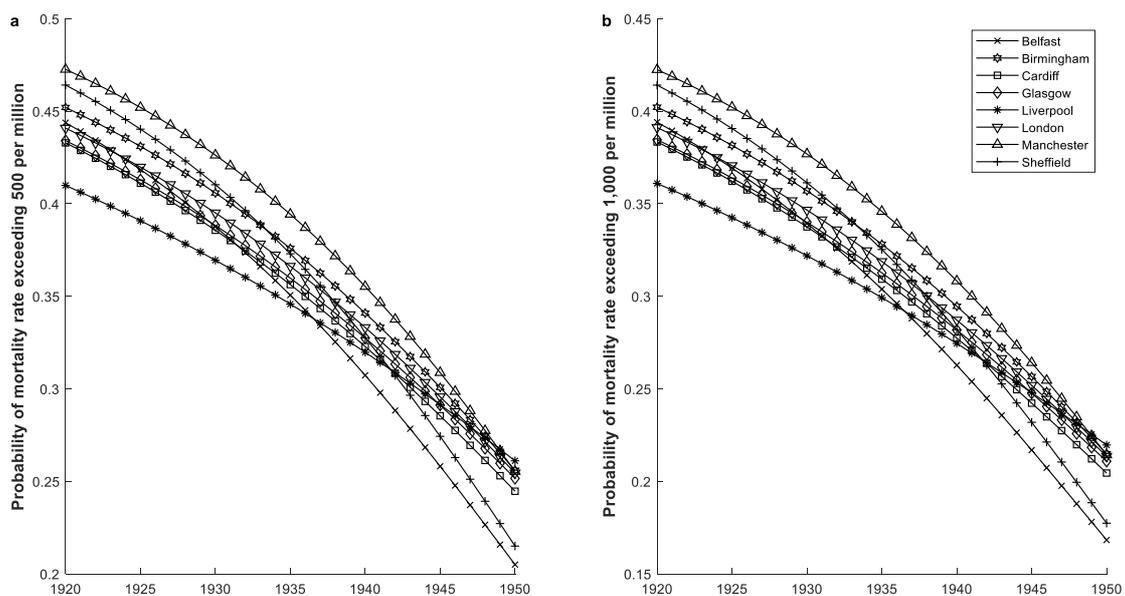
disease outbreak remains high until the 1940s, the dynamic pattern is different from that in Figure 3, implying disease outbreak risk that is very high initially and declines more rapidly. Overall, agnostic modelling of d_l and d_u confirms high and persistent mortality risk after the main pandemic waves, as well as similarity across geographical units, but, compared with the post-pandemic experience, it exaggerates it relative to the actual experience. On the other hand, exploiting the information pertinent to the pandemic in parameterising d_l and d_u leads to more accurate mortality risk predictions.

Figure S5: Model predicted influenza mortality rates following the 1918-19 pandemic (model with theoretical bounds)



Simulated median (solid black line), interquartile range (dark shading) and 80% prediction interval (light shading) are based on 1m random draws. Simulated outcomes are based on the model fitted using the average across cities. Upper and lower bounds are set to 1,000,000 and 1 respectively. Data are overplotted in red. Data for UK cities are taken from Medical Officer for Health reports.

Figure S6: Outbreak risk following the 1918-19 pandemic (model with theoretical bounds)



Outbreak probabilities computed from models fitted to each city. Upper and lower bounds are set to 1,000,000 and 1 respectively.

Next, we discuss the assumption that mortality rates after the main pandemic waves are determined by bounded Pareto distributions. The bounded Pareto distribution has been used in modelling outbreak fatality risk²¹ and offers flexibility and tractability in capturing important properties of a distribution of mortality rates in the context of our analysis. A model of the dynamics of mortality risk after the main pandemic waves needs to capture mortality risk in a period spanning high mortality rates, immediately after the main waves of the pandemic, and low mortality rates a few decades after the pandemic when the effect of the pandemic has died out. Hence it is important that the underlying mortality distribution in any given year can simultaneously account for a fat tail, and thus a potentially higher probability of disease outbreaks, and ensure that the mass of the distribution remains at the lower end of mortality. Over time, the tail becomes less important, tending to background mortality, and the distribution converges to a very high concentration toward the lower bound of mortality rates. The bounded Pareto distribution has the flexibility to account in a tractable manner for both characteristics: it has a fat tail while the most likely outcomes remain near the lower bound of mortality; and, conditional on the bounds, a transition in the tail probabilities is identified via changes in one parameter.

To further illustrate key points of the relevance of the Pareto distribution, we describe and fit a model with an alternative one-parameter distribution that allows for high probabilities for outcomes associated with the tail, the one-parameter Weibull-type distribution.³³ In this case, mortality rates are drawn from:

$$d_t \sim W_t(w_t), \quad (4)$$

where $w_t \in [0,1]$, noting that the tail contracts as w_t decreases. Assume that:

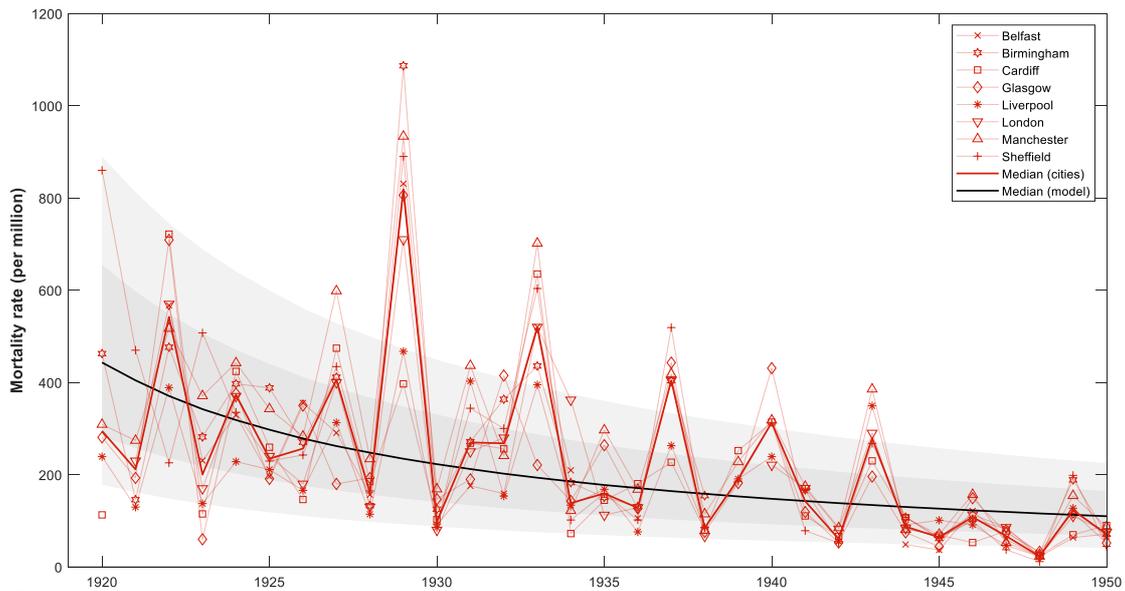
$$w_t = w_0 e^{-\lambda t}. \quad (5)$$

Conditional on the time process in (5), and thus conditional on the sequence $(w_t)_{t=0}^N$, d_t is independently distributed over time following (4). The likelihood is given by:

$$L(\lambda, w_0) = \prod_{t=0}^N \frac{\log^3(w_0 e^{-\lambda t})}{\log(w_0 e^{-\lambda t}) - 2} \bar{d}_t (\bar{d}_t + 1) (w_0 e^{-\lambda t})^{\bar{d}_t},$$

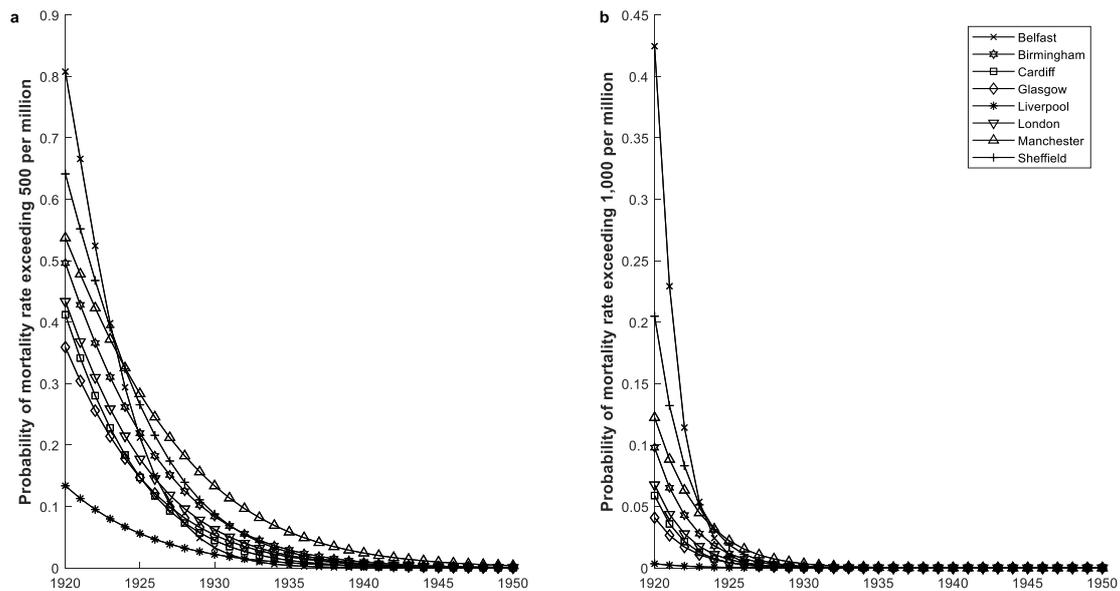
for a sample of mortality rates $(\bar{d}_t)_{t=0}^N$. Figures S7 and S8 reproduce Figures 2 and 3 under this modelling. The results from this model also reveal that mortality risk remains high for a long period after the main pandemic waves, and that its dynamic pattern is similar across geographical units. However, as can be seen, the predicted probabilities for disease outbreaks are higher than those in Figure 3. This is an implication of the one-parameter Weibull form that delivers a fat tail by shifting the mass of the distribution away from lower numbers.³³ Moreover, disease outbreak risk inherits a rate of rapid decline from the exponential decay of w_t , the Weibull parameter that determines the thickness of the tail.

Figure S7: Model predicted influenza mortality rates following the 1918-19 pandemic (One-parameter Weibull model)



Simulated median (solid black line), interquartile range (dark shading) and 80% prediction interval (light shading) are based on 1m random draws. Simulated outcomes are based on the model assuming a sequence of one-parameter Weibull distributions fitted using the average across cities (1920-1950). Data are overplotted in red. Data for UK cities are taken from Medical Officer for Health reports.

Figure S8: Outbreak risk following the 1918-19 pandemic (One-parameter Weibull model)



Outbreak probabilities computed from models fitted to each city assuming a sequence of one-parameter Weibull distributions.

The analysis above illustrates the importance of the property of the Pareto distribution that it can accommodate a fat tail with the mass of the distribution near the lower bound of outcomes. Alternative distributions that can deliver a concentration at lower mortality levels while also allowing for high probabilities associated with tail outcomes require more parameters to be specified (e.g. lognormal, Gaussian mixture) and require more assumptions regarding the dynamic transition. In particular, the model must specify the dynamic evolution of two or more parameters and a means to identify the specific combination of the dynamic processes of these parameters that characterises the evolution of

tail probabilities and of mortality risk more generally. Data availability restricts these options. Being a one-parameter distribution conditional on the bounds for mortality rates, the bounded Pareto offers a transparent way to model the post-pandemic dynamic evolution of mortality and disease outbreak risk.

Mortality risk predictions following COVID-19 under model uncertainty

We are interested in making predictions of probabilities of disease outbreaks at the point in time towards the end of the main waves of a pandemic, when there is information on the mortality of the main waves of the pandemic, and an expectation about the level of background mortality that should be reached after some time. The generality of the main qualitative characteristics of the dynamic risk patterns in Figures 2 and 3, across different geographical units with different experience of main wave and background, suggests that the model in (1) – (2) can be used for counterfactual analysis. Nevertheless, the specific quantitative predictions may depend on the specific sequence $(\alpha_t)_{t=0}^N$ used, or equivalently on the specific parameters η_0 and λ . Inspection of Table S2 suggests that even though these parameters are relatively similar, they still vary across geographical units. This variation generates model uncertainty regarding the risk projections, because it creates uncertainty about which sequence of distributions in $(PD_t)_{t=0}^N$ generate the data, which is different from the epidemiological uncertainty regarding mortality rates that is generated by a given sequence of distributions.

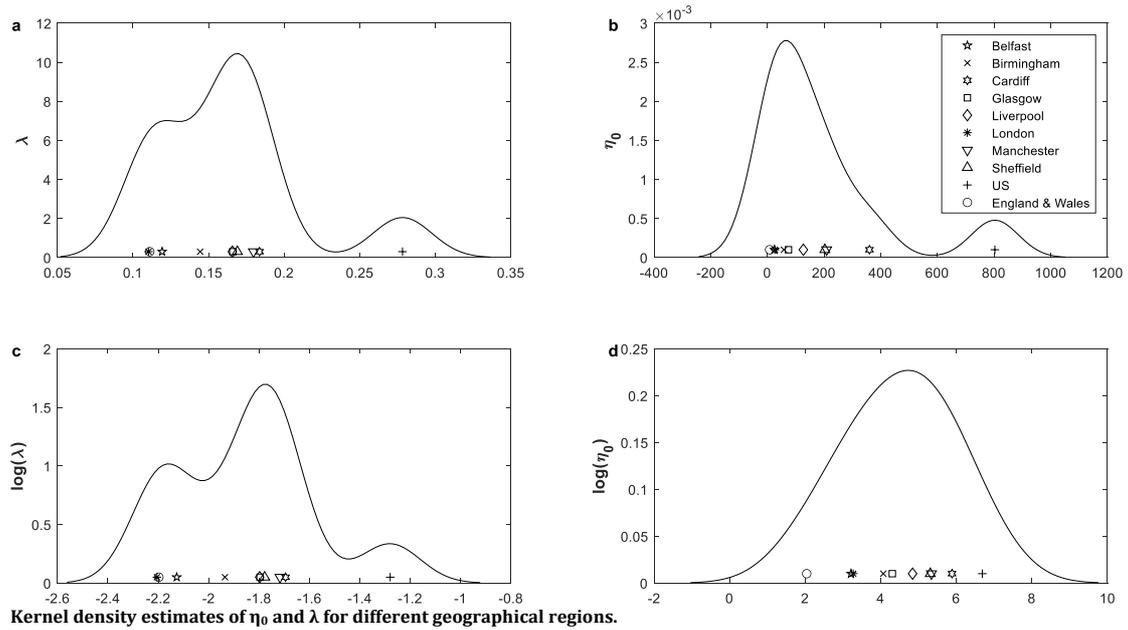
To account for such model uncertainty, we assume that the parameters η_0 and λ that are relevant to a new pandemic are drawn from the same distribution from which the parameters in Table S2 are drawn, and then perform a Monte Carlo analysis that provides a distribution of possible outcomes as a function of draws of η_0 and λ from that distribution. Using the parameter values of η_0 and λ in Table S2, we estimate the underlying distribution (see Figure S9) and approximate it as a joint lognormal:

$$\begin{pmatrix} \log(\eta_0) \\ \log(\lambda) \end{pmatrix} \sim MVN(\mu, \sigma)$$

$$\mu = \begin{pmatrix} -1.85 \\ 4.5 \end{pmatrix}, \sigma = \begin{pmatrix} 0.08 & 0.37 \\ 0.37 & 1.96 \end{pmatrix}$$

We then draw one million pairs of η_0 and λ from the implied joint density and summarise relevant percentiles of the generated distribution of predicted probabilities of mortality rates in Figure 4.

Figure S9: Distribution of η_0 and λ implied by the models fitted to previous pandemics



Supplementary Material C

Additional Figures

Figure S10: Model predicted influenza mortality rates following the 1918-19 and 1890-91 pandemics (deaths per million)

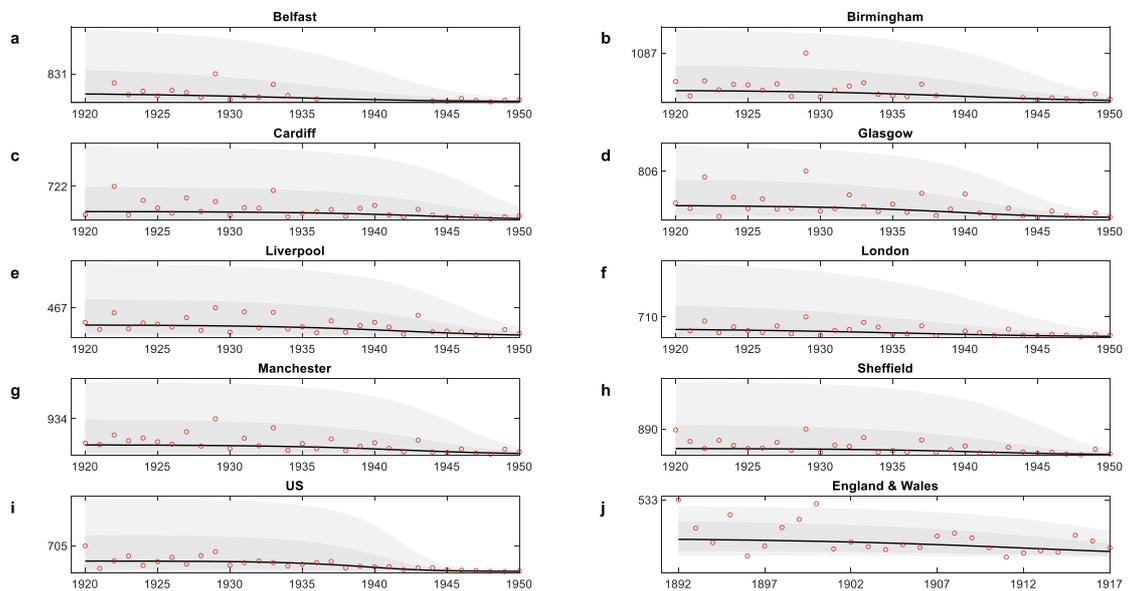
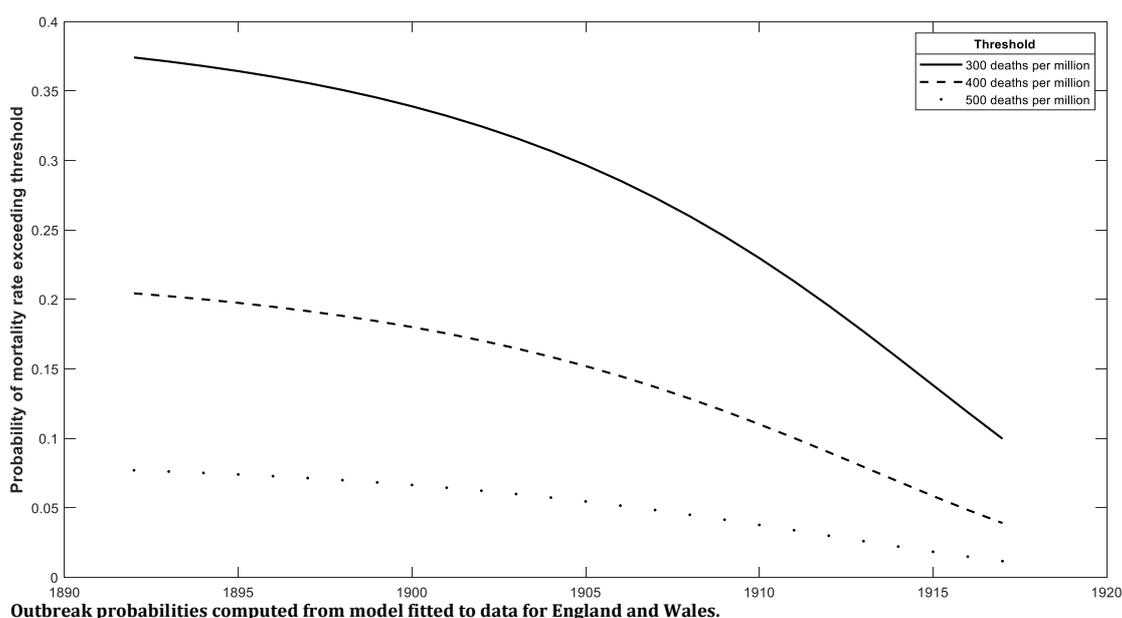


Figure S11: Outbreak risk after the 1890-91 pandemic



Conflicts of interest

The authors declare no conflicts of interest.

Data availability

The municipal level dataset generated during the current study are available on GitHub at <https://github.com/maxschr90/Schroeder-et-al.-2021--How-long-do-pandemics-last->. Data for the US and for England and Wales are available from original publications, as described and referenced in the caption of Figure 1.

Code availability

Code is freely available on GitHub at <https://github.com/maxschr90/Schroeder-et-al.-2021--How-long-do-pandemics-last->.

Acknowledgements and funding

We would like to thank Irene O'Brien at Glasgow City Archives for guidance on archival data, Rosie Sida for help with data preparation, and Dan Haydon, Sema Nickbakhsh, Richard Reeve and Denise Brown for comments on an earlier version of this work.

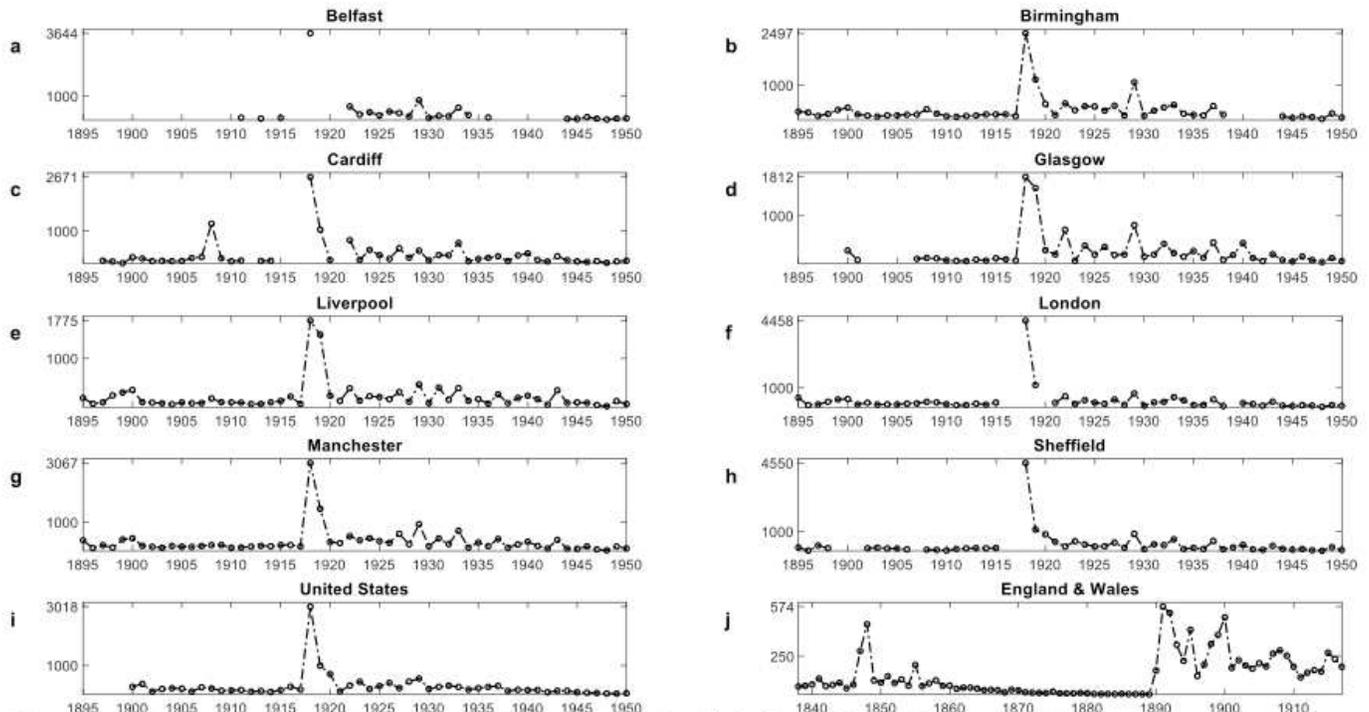
This project is supported by a grant that is funded by the Economic and Social Research Council (ESRC) as part of UK Research and Innovation's rapid response to COVID-19 (ES/V005898/1). The project also received support from the College of Social Sciences College Research Fund 2020/21, and from the Adam Smith Business School Student Internship Scheme, University of Glasgow. The project is supported by Glasgow City Archives, Glasgow Life. Rebecca Mancy is supported by The Leckie Fellowship, the UK Medical Research Council (MRC) Places and Health Programme (MC_UU_00022/4) and the Chief Scientist Office (CSO) (SPHSU10) at the MRC/CSO Social and Public Health Sciences Unit, University of Glasgow.

References

1. Marmot, M. Health equity in England: The Marmot review 10 years on. *The BMJ* **368**, m693 (2020).
2. Marmot, M. & Allen, J. COVID-19: exposing and amplifying inequalities. *Journal of Epidemiology and Community Health* **74**, 681–682 (2020).
3. Stantcheva, S. Inequalities in the Times of a Pandemic. *Economic Policy* (2021).
4. Madhav, N. *et al.* Pandemics: Risks, Impacts, and Mitigation. in *Disease Control Priorities: Improving Health and Reducing Poverty. 3rd edition* (The International Bank for Reconstruction and Development/The World Bank, 2017).
5. Scheidel, W. *The great Leveler*. (Princeton University Press, 2018).
6. Lavine, J. S., Bjornstad, O. N. & Antia, R. Immunological characteristics govern the transition of COVID-19 to endemicity. *Science* **371**, 741–745 (2021).
7. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
8. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control*. (Oxford University Press, 1992).
9. Oxford, J. S. Influenza A pandemics of the 20th century with special reference to 1918: virology, pathology and epidemiology. *Reviews in Medical Virology* **10**, 119–133 (2000).
10. Torjesen, I. Covid-19 will become endemic but with decreased potency over time, scientists believe. *BMJ* **372**, n494 (2021).
11. McKie, R. UK can expect thousands of Covid deaths every year, warn scientists. *the Guardian*
<https://www.theguardian.com/world/2021/jul/31/uk-can-expect-thousands-of-covid-deaths-every-year-warn-scientists> (2021).
12. Phillips, N. The coronavirus is here to stay - here's what that means. *Nature* **590**, 382–384 (2021).
13. Quinn, S. C. & Kumar, S. Health inequalities and infectious disease epidemics: A challenge for global health security. *Biosecurity and Bioterrorism* **12**, 263–273 (2014).
14. Mamelund, S.-E. & Dimka, J. Social inequalities in infectious diseases. *Scandinavian Journal of Public Health* 140349482199722 (2021) doi:10.1177/1403494821997228.
15. Angelopoulos, K., Lazarakis, S., Mancy, R. & Schroeder, M. *Post-pandemic wealth and health inequality under recurrent outbreak risk*. (2021).
16. Linder, F. E. & Grove, R. D. *Vital statistics rates in the United States, 1900-1940*. (US Government Printing Office, 1943).
17. Grove, R. D. & Hetzel, A. M. *Vital statistics rates in the United States, 1940-1960*. vol. 83 (US Department of Health, Education, and Welfare, Public Health Service ..., 1968).

18. Langford, C. The age pattern of mortality in the 1918-19 influenza pandemic: an attempted explanation based on data for England and Wales. *Medical History* **46**, 1–20 (2002).
19. Charles River Editors. *The 1889-1890 Flu Pandemic: The History of the 19th Century's Last Major Global Outbreak*. (2020).
20. Beveridge, W. I. The chronicle of influenza epidemics. *History and philosophy of the life sciences* **13**, 223–234 (1991).
21. Cirillo, P. & Taleb, N. N. Tail risk of contagious diseases. *Nature Physics* **2020 16:6 16**, 606–613 (2020).
22. Bloom, N. The Impact of Uncertainty Shocks. *Econometrica* **77**, 623–685 (2009).
23. Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I. & Terry, S. J. Really Uncertain Business Cycles. *Econometrica* **86**, 1031–1065 (2018).
24. Heathcote, J., Storesletten, K. & Violante, G. L. The macroeconomic implications of rising wage inequality in the United States. *J. Polit. Econ.* **118**, 681–722 (2010).
25. Angelopoulos, K., Lazarakis, S. & Malley, J. The distributional implications of asymmetric income dynamics. *European Economic Review* **128**, 103502 (2020).
26. Angelopoulos, K., Lazarakis, S. & Malley, J. Cyclical labour income risk in Great Britain. *Journal of Applied Econometrics* (2021) doi:10.1002/JAE.2860.
27. Wellcome Trust. The Medical Officer of Health reports. <https://wellcomelibrary.org/moh/about-the-reports/about-the-medical-officer-of-health-reports/> (2021).
28. Vital Statistics of the US - Homepage. <https://www.cdc.gov/nchs/products/vsus.htm> (2015).
29. Ritchie, H. *et al.* Coronavirus Pandemic (COVID-19). *Published online at OurWorldInData.org* <https://ourworldindata.org/coronavirus/country/united-kingdom#what-is-the-cumulative-number-of-confirmed-deaths> (2020).
30. ONS. Population estimates - Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates> (2021).
31. ONS. Influenza deaths in 2018, 2019 and 2020 - Office for National Statistics. <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/influenzadeathsin20182019and2020> (2021).
32. ONS. Population estimates for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2018> (2019).
33. Alexopoulos, A. One-Parameter Weibull-Type Distribution, Its Relative Entropy with Respect to Weibull and a Fractional Two-Parameter Exponential Distribution. *Stats* **2**, 34–54 (2019).

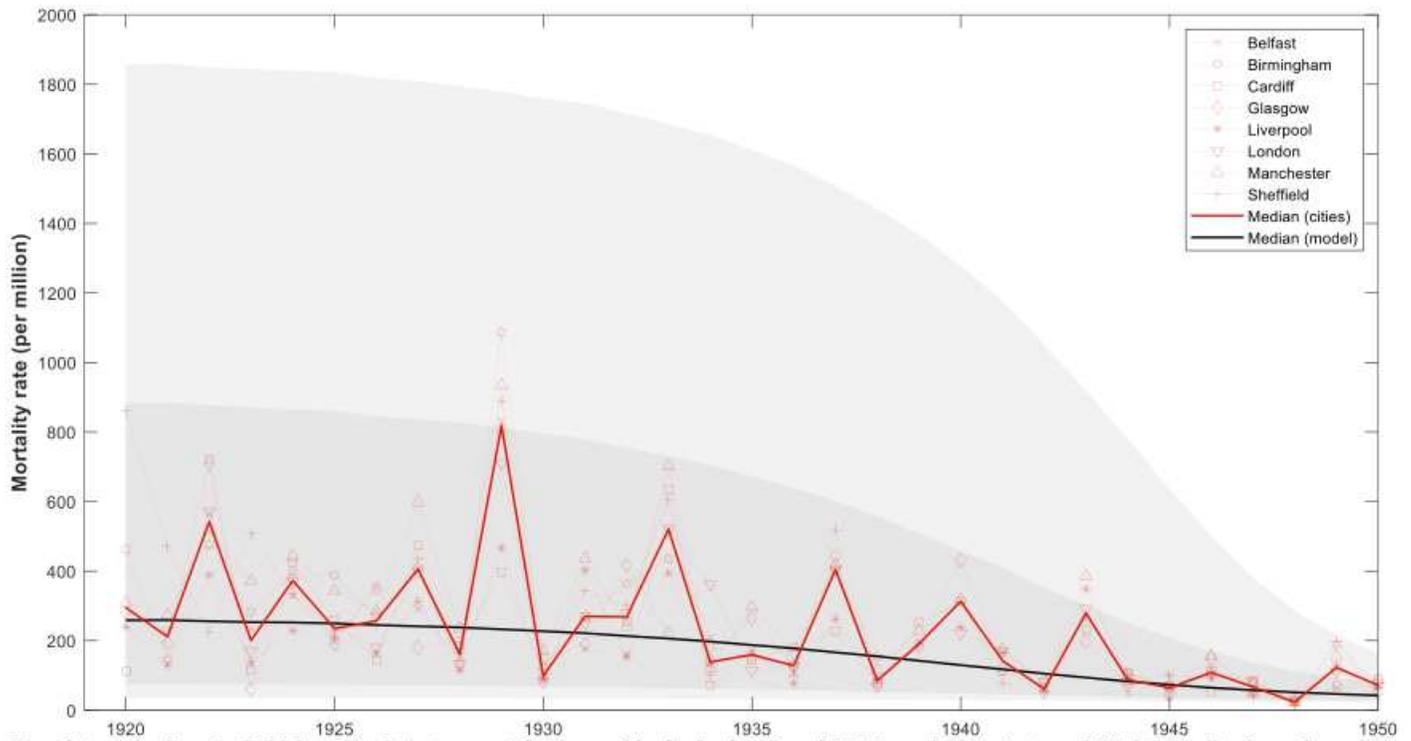
Figures



Mortality rate in deaths per million population. The upper y-axis tick indicates the maximum mortality rate for each data series. Data for UK cities are taken from Medical Officer for Health reports. Data for US is taken from Lindner and Grove (1943)¹⁶ and Grove and Hetzel (1968).¹⁷ Data for England & Wales are taken from Langford (2002),¹⁸ Table 5. Information on population sizes and densities is provided in Table S2.

Figure 1

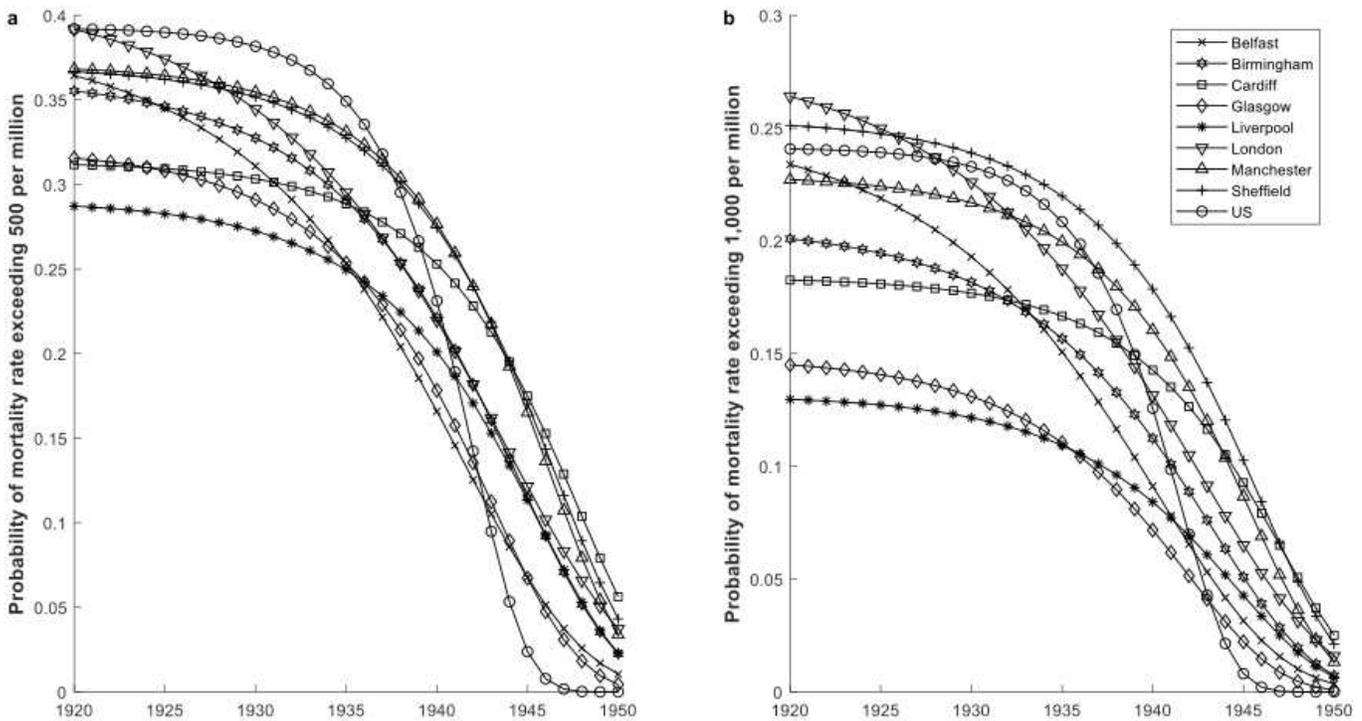
Annual mortality rate from influenza



Simulated median (solid black line), interquartile range (dark shading) and 80% prediction interval (light shading) are based on 1m random draws. Simulated outcomes are based on the model fitted using the average mortality rates across cities. Data are overlaid in red. Data for UK cities are taken from Medical Officer for Health reports.

Figure 2

Model predicted influenza mortality rates following the 1918-19 pandemic



Outbreak probabilities computed from models fitted to each city and the US, for different thresholds

Figure 3

Outbreak risk following the 1918-19 pandemic

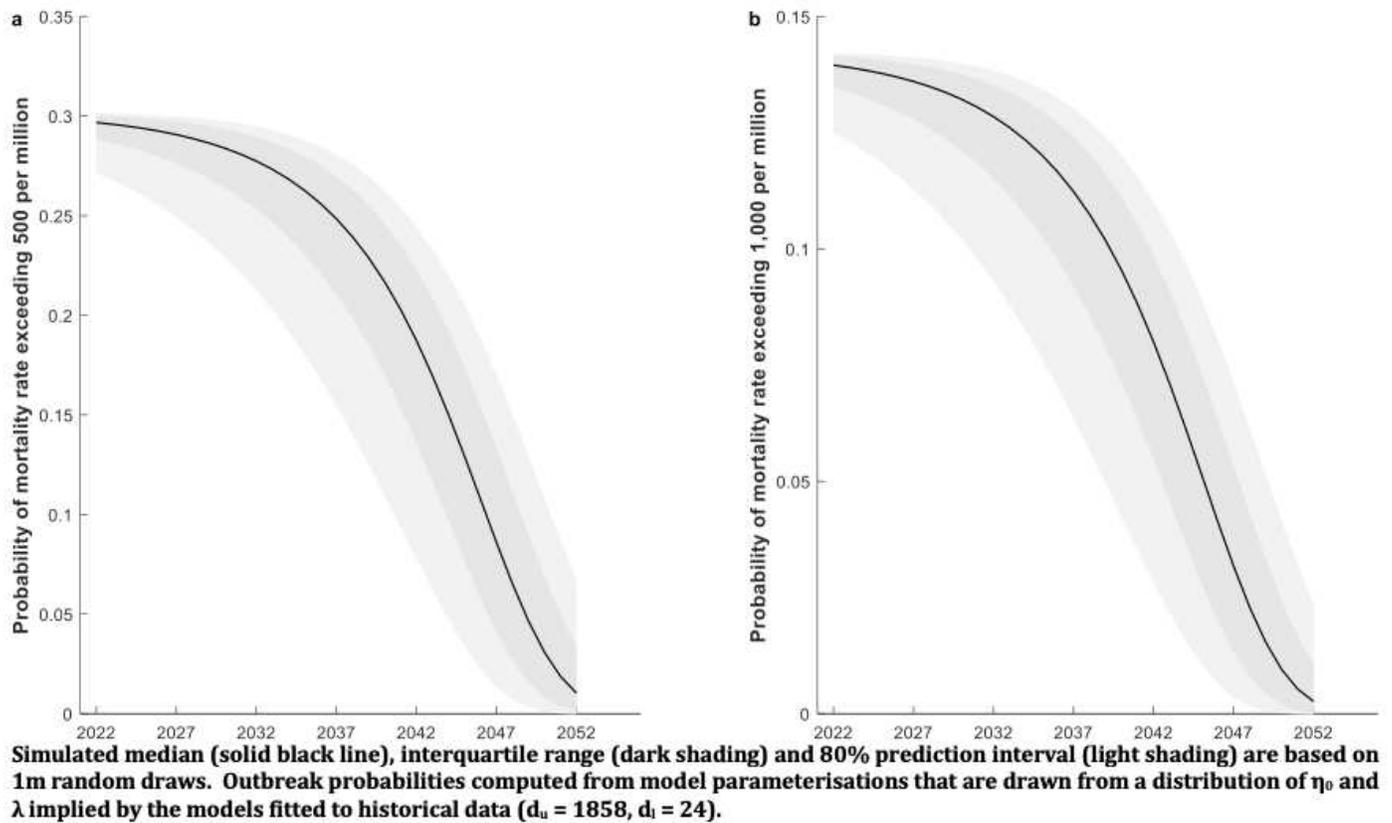


Figure 4

Predicted outbreak risk following the 2020-21 pandemic

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)