

# Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study

Amitava Banerjee, Laura Pasea, Steve Harris, Arturo Gonzalez-Izquierdo, Ana Torralbo, Laura Shallcross, Mahdad Noursadeghi, Deenan Pillay, Neil Sebire, Chris Holmes, Christina Pagel, Wai Keong Wong, Claudia Langenberg, Bryan Williams, Spiros Denaxas, Harry Hemingway



## Summary

**Background** The medical, societal, and economic impact of the coronavirus disease 2019 (COVID-19) pandemic has unknown effects on overall population mortality. Previous models of population mortality are based on death over days among infected people, nearly all of whom thus far have underlying conditions. Models have not incorporated information on high-risk conditions or their longer-term baseline (pre-COVID-19) mortality. We estimated the excess number of deaths over 1 year under different COVID-19 incidence scenarios based on varying levels of transmission suppression and differing mortality impacts based on different relative risks for the disease.

**Methods** In this population-based cohort study, we used linked primary and secondary care electronic health records from England (Health Data Research UK–CALIBER). We report prevalence of underlying conditions defined by Public Health England guidelines (from March 16, 2020) in individuals aged 30 years or older registered with a practice between 1997 and 2017, using validated, openly available phenotypes for each condition. We estimated 1-year mortality in each condition, developing simple models (and a tool for calculation) of excess COVID-19-related deaths, assuming relative impact (as relative risks [RRs]) of the COVID-19 pandemic (compared with background mortality) of 1·5, 2·0, and 3·0 at differing infection rate scenarios, including full suppression (0·001%), partial suppression (1%), mitigation (10%), and do nothing (80%). We also developed an online, public, prototype risk calculator for excess death estimation.

**Findings** We included 3 862 012 individuals (1 957 935 [50·7%] women and 1 904 077 [49·3%] men). We estimated that more than 20% of the study population are in the high-risk category, of whom 13·7% were older than 70 years and 6·3% were aged 70 years or younger with at least one underlying condition. 1-year mortality in the high-risk population was estimated to be 4·46% (95% CI 4·41–4·51). Age and underlying conditions combined to influence background risk, varying markedly across conditions. In a full suppression scenario in the UK population, we estimated that there would be two excess deaths (*vs* baseline deaths) with an RR of 1·5, four with an RR of 2·0, and seven with an RR of 3·0. In a mitigation scenario, we estimated 18 374 excess deaths with an RR of 1·5, 36 749 with an RR of 2·0, and 73 498 with an RR of 3·0. In a do nothing scenario, we estimated 146 996 excess deaths with an RR of 1·5, 293 991 with an RR of 2·0, and 587 982 with an RR of 3·0.

**Interpretation** We provide policy makers, researchers, and the public a simple model and an online tool for understanding excess mortality over 1 year from the COVID-19 pandemic, based on age, sex, and underlying condition-specific estimates. These results signal the need for sustained stringent suppression measures as well as sustained efforts to target those at highest risk because of underlying conditions with a range of preventive interventions. Countries should assess the overall (direct and indirect) effects of the pandemic on excess mortality.

**Funding** National Institute for Health Research University College London Hospitals Biomedical Research Centre, Health Data Research UK.

**Copyright** © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

## Introduction

Excess deaths from the coronavirus disease 2019 (COVID-19) pandemic might arise both in those infected (direct effects), as well as those affected (indirectly, not infected) by altered access to health services; the physical, psychological, and social effects of distancing; and economic changes. Understanding the effect of COVID-19 on mortality during this emergency requires modelling of an infectious disease, as well as wider medical and

societal changes. One way of estimating and monitoring excess mortality is to compare observed numbers of deaths with those expected based on the background (pre-COVID-19) mortality risks in the population.<sup>1</sup>

One model of the population mortality impact of COVID-19 is based on age-stratified death rates over days in infected patients, but excludes prevalence of underlying conditions, their differing pre-COVID-19 background long-term mortality risks, or the additional risk associated with

*Lancet* 2020; 395: 1715–25

Published Online

May 12, 2020

[https://doi.org/10.1016/S0140-6736\(20\)30854-0](https://doi.org/10.1016/S0140-6736(20)30854-0)

**Institute of Health Informatics** (A Banerjee DPhil, L Pasea PhD, A Gonzalez-Izquierdo PhD, A Torralbo PhD, L Shallcross PhD, Prof S Denaxas PhD, Prof H Hemingway FMedSci), **Division of Infection and Immunity** (Prof M Noursadeghi PhD, Prof D Pillay PhD), **Clinical Operational Research Unit** (Prof C Pagel PhD), and **Institute of Cardiovascular Science** (Prof B Williams PhD), **University College London, London, UK**; **University College London Hospitals NHS Trust, London, UK** (A Banerjee, S Harris PhD, W K Wong PhD, Prof B Williams); **Barts Health NHS Trust, The Royal London Hospital, London, UK** (A Banerjee); **Great Ormond Street Hospital, London, UK** (Prof N Sebire PhD); **University of Oxford, Oxford, UK** (Prof C Holmes PhD); **Alan Turing Institute, London, UK** (Prof C Holmes, Prof S Denaxas); **MRC Epidemiology Unit, University of Cambridge, Cambridge, UK** (C Langenberg PhD); **University College London Hospitals National Institute for Health Research Biomedical Research Centre, London, UK** (Prof B Williams); and **Health Data Research UK, London, UK** (Prof S Denaxas, Prof H Hemingway)

Correspondence to:

Dr Amitava Banerjee, Institute of Health Informatics, University College London, London NW1 2DA, UK  
[ami.banerjee@ucl.ac.uk](mailto:ami.banerjee@ucl.ac.uk)

### Research in context

#### Evidence before this study

We searched PubMed, medRxiv, bioRxiv, arXiv, and Wellcome Open Research for peer-reviewed articles, preprints, and research reports on mortality and comorbidities at baseline in coronavirus disease 2019 (COVID-19), using the search terms “coronavirus”, “COVID-19”, and similar terms, and “mortality”, up to March 21, 2020. We found no previous studies of excess deaths due to the COVID-19 pandemic. Previous tools and models to estimate COVID-19-related deaths have not considered the prevalence of underlying conditions, background pre-COVID-19 mortality, long-term estimates, or the relative impact (relative risk) of COVID-19 on mortality. Without these data, it is challenging to predict and prepare health systems to respond to the COVID-19 health-care burden.

#### Added value of this study

We produced a model to estimate the excess COVID-19-related mortality by incorporating population infection rate in different scenarios relating to degree of social isolation, differing degrees

of impact of COVID-19 on health systems, and prevalence of underlying conditions. We showed that in a full suppression scenario, even if COVID-19 has high relative impact on health systems, excess mortality over 1 year can be minimised. Conversely, with mitigation (ie, less rigorous and voluntary measures), we predict between 18 000 and 37 000 deaths. We have provided a prototype risk calculator for the researchers and policy makers to make estimates and change the assumptions in the model. We also highlight the value of routine electronic health records and use of large, linked datasets at the country level in responding to the pandemic.

#### Implications of all the available evidence

Our models allow the estimation of overall (direct and indirect) effects on excess mortality of the COVID-19 pandemic. For the first time, we provide a model, a tool, and open-access data to individualise risk prediction across underlying conditions, which has clinical, public health, and research benefits for both the COVID-19 pandemic and post-pandemic contexts.

COVID-19.<sup>2</sup> Few reports of excess deaths beyond specific high-risk populations have been published<sup>3,4</sup> (most deaths have occurred in people with underlying health conditions or those of older ages<sup>5-7</sup>). This situation is changing, with severe infections being treated in younger patients with COVID-19 who do not have underlying conditions.<sup>8</sup> Case fatality rates for COVID-19 vary from 0.27% to 10%,<sup>9</sup> possibly explained by differing demography, testing strategies, and prevalence of underlying conditions. The UK has relatively high case fatality rates (8.2%), but mortality rates are unknown at this stage of the pandemic because testing is more common among sicker patients who are admitted to hospital (the context where most testing has been done) rather than milder cases. On April 14, the Office for National Statistics reported 6000 excess deaths registered in the week March 28 to April 3, 2020, of which about 2500 deaths did not have COVID-19 recorded on the death certificates, providing the first indication of indirect effects of the pandemic on mortality.<sup>10</sup>

Physical distancing and other strategies focus on high-risk groups, with US Centers for Disease Control and Prevention specifying “older adults, and people who have serious chronic medical conditions such as heart disease, diabetes and lung disease”.<sup>11</sup> On March 16, 2020, the UK Government announced that particular subgroups are at high risk from COVID-19,<sup>12</sup> and recommended stringent distancing measures, telling people how to stay away from others. On March 22, 2020, the Government announced that 1.5 million “extremely vulnerable” people in England with underlying conditions (including severe chronic obstructive pulmonary disease [COPD]) should “shield” for 12 weeks, but did not explain how these conditions were selected.<sup>13</sup> The UK Government did not implement full suppression (during which only key workers are likely

to be exposed) until March 23, 2020. Up to that date, people were still in close proximity in work and public places.

Here, we provide estimates of 1-year mortality by underlying conditions to understand the broader health impact of COVID-19. Our objectives were to provide the research and policy community and public with parameters (prevalence and background pre-COVID-19 1-year mortality risk by age and underlying conditions) to assist modelling; and to provide initial estimates of the excess COVID-19-related deaths over a 1-year period based on differing rates and relative levels of impact of infection. We define COVID-19-related mortality as both direct and indirect effects of the pandemic.

## Methods

### Study design and data sources

This was a population-based cohort study, done with clinical research using linked bespoke studies and electronic health records (CALIBER), an open research platform with validated, reusable definitions of several hundred underlying conditions.<sup>14,15</sup> It links electronic health records from different data sources (via UK unique individual identification data and National Health Service [NHS] numbers), including primary care (Clinical Practice Research Datalink-GOLD), hospital care (Hospital Episodes Statistics), and death registry (Office of National Statistics). Nearly all (>99%) of the English population has general practice registration. CALIBER has been shown to be representative of the general population in terms of sociodemographic characteristics and overall mortality. Approval for the study was granted by the Independent Scientific Advisory Committee (20\_074R) of the Medicines and Healthcare products Regulatory Agency in the UK in accordance with the Declaration of Helsinki.

## Study population

Eligible individuals were aged 30 years or older and registered with a general practice at any point between Jan 1, 1997, and Jan 1, 2017, with at least 1 year of follow-up data. Demographic characteristics (age, gender, index of multiple deprivation quintiles, and geographical region) and baseline characteristics (prevalence of underlying conditions, risk factors, and medications) were recorded.<sup>14</sup> The overall population at high risk of COVID-19 and recommended for physical distancing was defined as per Public Health England guidance.<sup>12</sup>

## Open national portal for definitions of underlying conditions

We used the Health Data Research (HDR) UK–CALIBER open online portal to define underlying conditions from records. The underlying conditions were defined using Read codes in primary care and the International Classification of Diseases 10th edition for hospital admissions as per validated CALIBER phenotypes.<sup>14</sup> Hypertension was defined on the basis of recorded values in primary care according to the most recent guidelines ( $\geq 140$  mm Hg systolic blood pressure [or  $\geq 150$  mm Hg for people aged  $\geq 60$  years without diabetes and chronic kidney disease] or  $\geq 90$  mm Hg diastolic blood pressure).<sup>16</sup> Diabetes was defined at baseline (including type 1, type 2, or uncertain type) on the basis of coded diagnoses recorded in Clinical Practice Research Datalink or Hospital Episode Statistics at or before study entry.<sup>15</sup> Severe obesity was defined as body-mass index of 40 kg/m<sup>2</sup> or more.<sup>16</sup> Cardiovascular disease was defined as the 12 most common symptomatic manifestations: chronic stable angina, unstable angina, myocardial infarction, unheralded death from coronary heart disease, heart failure, cardiac arrest or sudden coronary death, transient ischaemic attack, ischaemic stroke, intracerebral haemorrhage, subarachnoid haemorrhage, peripheral arterial disease, and abdominal aortic aneurysm, as per previous studies.<sup>15–17</sup> COPD and chronic kidney disease were defined using previous definitions.<sup>14–17</sup> Multimorbidity was defined as the co-occurrence of two or more conditions in an individual.<sup>18</sup> Given recent interest in the potential roles of angiotensin-converting enzyme (ACE) inhibitors<sup>19</sup> and non-steroidal inflammatory drugs (NSAIDs),<sup>20</sup> we also estimated prevalence of their use.

## Statistical analysis

To assess prevalent underlying conditions, we only included records from the year before baseline (date of general practice registration or being aged  $>30$  years during the 1997–2017 study period). Follow-up ceased at the date of death or on Jan 1, 2017. We estimated prevalence of each underlying condition and used Kaplan-Meier estimates of 1-year all-cause mortality at each age group cell to estimate the number of excess deaths by age bands and at each number of conditions cell to estimate by

number of underlying conditions. We then modelled the excess COVID-19-associated mortality for relative impact of mortality associated with COVID-19 (relative risk [RR] compared with background mortality) of 1·5, 2·0, and 3·0<sup>21</sup> at the infection rate consequences of different strategies: full suppression (0·001%), partial suppression (1%), mitigation (10%), or do nothing (80%).<sup>22</sup> Relative risk is a comparison of risks or probabilities. For example, an RR of 2·0 for COVID-19 describes a doubling of the risk of mortality in individuals infected with or affected by COVID-19 versus their background risk of death. A definition of COVID-19-related mortality relevant to the emergency includes both deaths among those infected (direct), and deaths among those affected by health service, societal, and economic changes (indirect). Considering excess deaths in this way, it is reasonable to consider that a high proportion (eg, 80% of the population) might be affected by the COVID-19 emergency. To extrapolate our estimates of excess deaths to the whole population of England older than 30 years, we used 2018 estimates of overall population size and mortality.<sup>23</sup> For illustration, we applied our study estimates to UN Population Fund<sup>24</sup> estimates of population size in other countries. All analyses were done using R, version 3.4.3. Data were analysed in the University College London ISO27001 safe haven.

## Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. LP and SD had full access to all the data in the study and AB and HH had final responsibility for the decision to submit for publication.

## Results

We included 3 862 012 individuals aged 30 years or older. 1 957 935 (50·7%) were women and 1 904 077 (49·3%)

For the HDR UK–CALIBER online portal see [www.caliberresearch.org/portal](http://www.caliberresearch.org/portal)

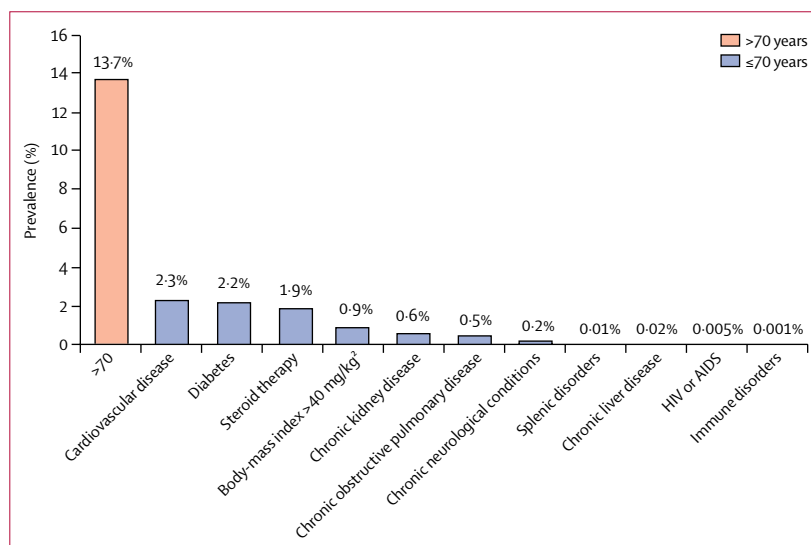


Figure 1: Prevalence of underlying conditions associated with high mortality in COVID-19 infection (n=3862012)

were men; 3 331 280 (86.3%) were aged 70 years or younger (mean age 43.5 years [SD 11.7] in both sexes) and 530 732 (13.7%) were older than 70 years (mean age 78.1 years [6.1] in men and 80.2 years [7.0] in women; appendix p 1).

More than 20.0% of the study population had at least one high-risk condition defined by Public Health England guidelines. 13.7% were older than 70 years and 6.3% were aged 70 years or younger with at least one underlying condition (figure 1). Age-specific (5-year age bands) and sex-specific estimates of prevalence for each underlying condition are available online.

Multimorbidity was common (10.1%), especially with cardiovascular disease, diabetes, chronic kidney disease, and COPD (appendix p 2). In individuals without cardiovascular disease (mean age 47.5 years [SD 16.2]), diabetes (48.0 years [16.5]), chronic kidney disease

(48.0 years [16.4]) and COPD (48.2 years [16.6]) 6.7% without cardiovascular disease, 7.9% without diabetes, 9.1% without chronic kidney disease, and 9.7% without COPD still had a condition meeting criteria for physical distancing. Hypertension was common: 7.4% in the individuals aged 70 years or younger and 31.7% in those older than 70 years. In individuals aged 70 years or younger, prescription of ACE inhibitors was lower (3.7%) than in those older than 70 years (14.9%). 14.0% of those aged 70 years or younger and 21.1% of those older than 70 years were prescribed NSAIDs.

Mortality at 1 year was estimated at 4.46% (95% CI 4.41–4.51) among individuals at high risk. 1-year mortality was 8.62% (8.32–8.92) for those with COPD, 7.84% (7.62–8.06) for those with chronic kidney disease, 6.37% (6.25–6.49) for those with cardiovascular disease, and 4.1% (4.0–4.2) for those with

See Online for appendix

For the open-access repository see <https://doi.org/10.6084/m9.figshare.12238118>

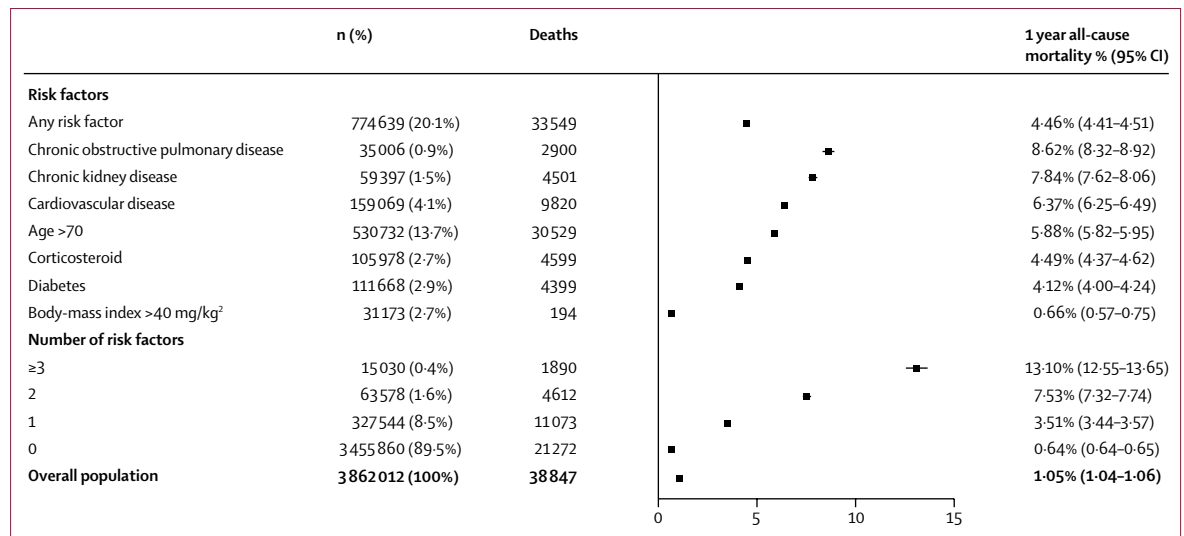


Figure 2: Baseline 1-year mortality in England according to underlying conditions (n=3 862 012)

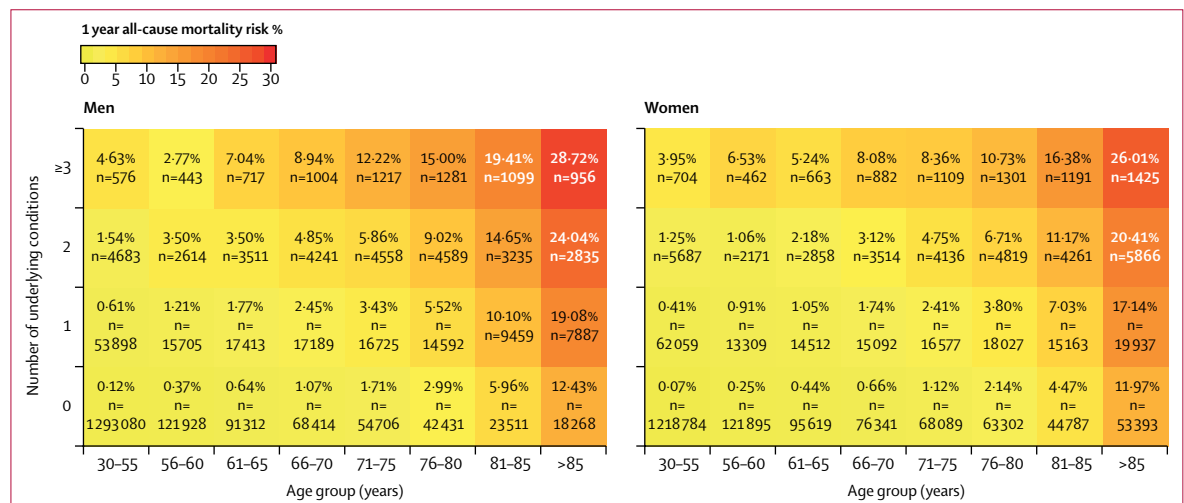
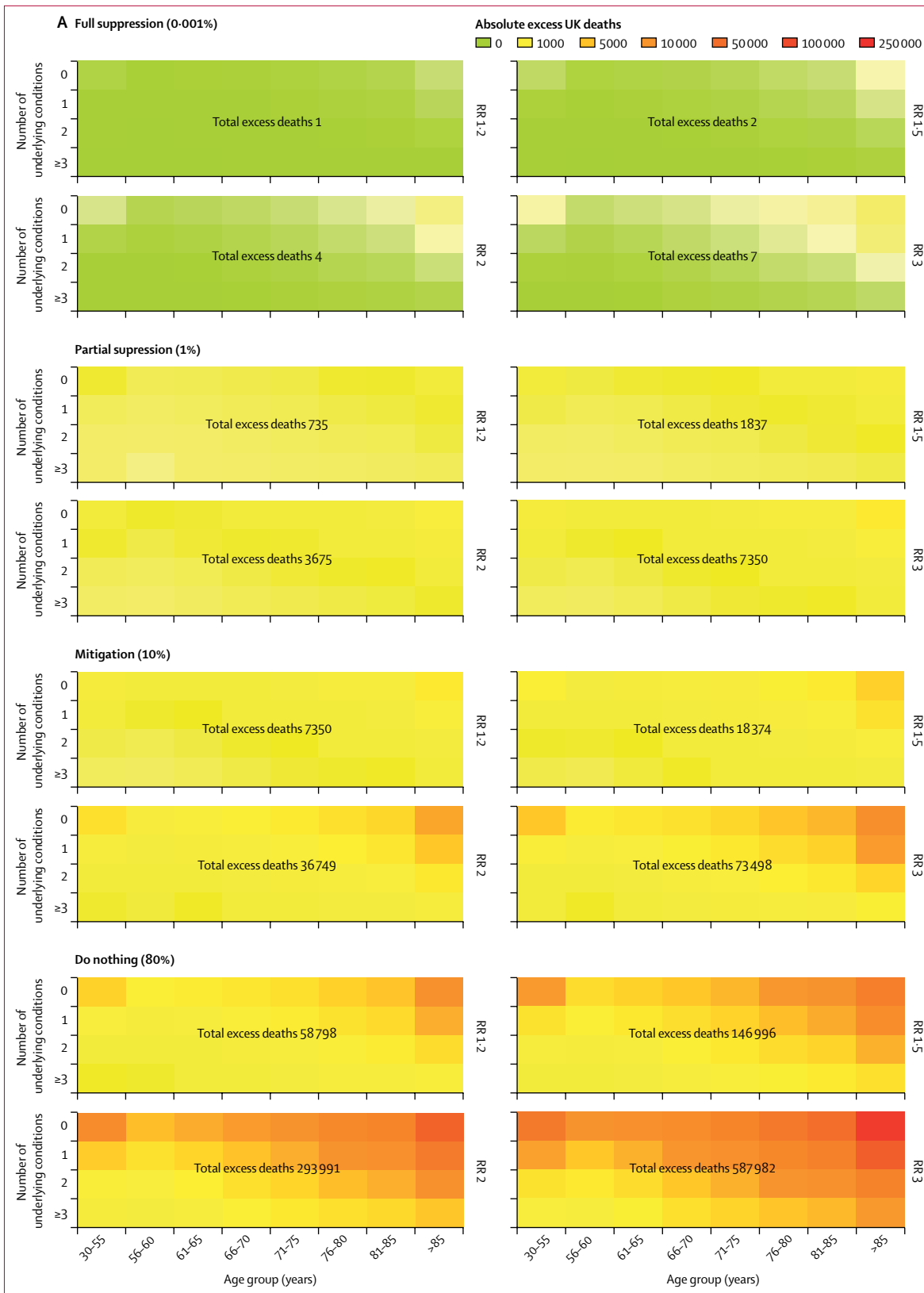
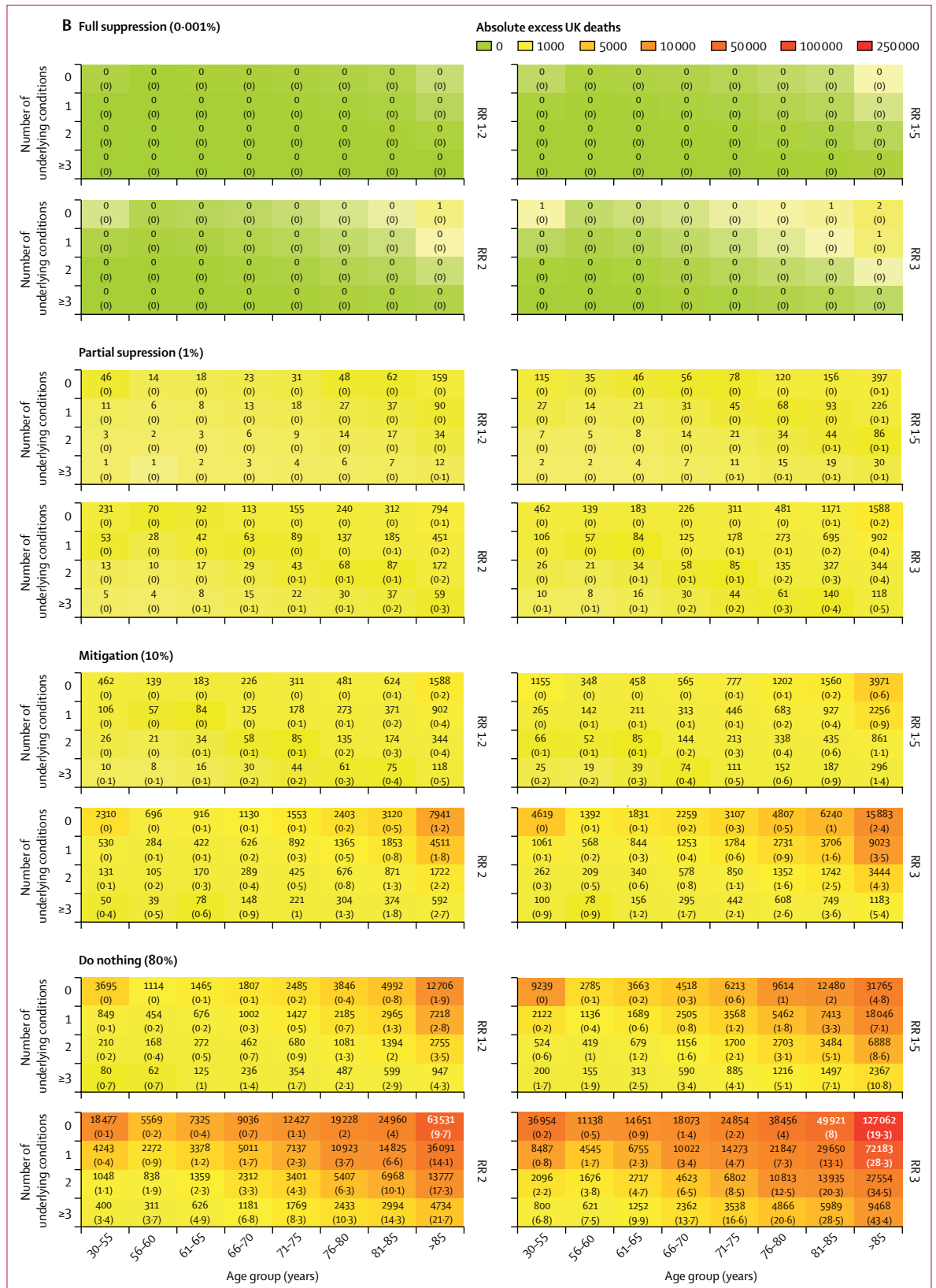


Figure 3: Baseline 1-year mortality in England according to number of underlying conditions, age category, and sex (n=3 862 012)



(Figure 4 continues on next page)



**Figure 4: Estimated number of excess deaths at 1 year in the UK at different infection rates and relative risks for the impact of COVID-19**  
 (A) Total deaths. (B) Detailed breakdown of deaths across different categories.  
 RR=relative risk.

diabetes (figure 2). Among those older than 70 years, 1-year mortality rates were 13.6% (12.9–14.3) for COPD, 11.5% (11.0–12.1) for chronic kidney disease, 10.4% (10.1–10.7) for cardiovascular disease, and 8.9% (8.5–9.4) for diabetes in men and 12.3% (11.6–13.0) for COPD, 9.6% (9.3–10.0) for chronic kidney disease, 10.6% (10.3–10.9) for cardiovascular disease, and 9.8% (9.4–10.2) for diabetes in women.

We estimated baseline 1-year mortality risk by age group, sex, and number of conditions, in the absence of COVID-19 (figure 3). For example, a man aged 66–70 years with no underlying high-risk conditions has a higher estimated 1-year mortality risk (1.07%) than a woman aged 56–60 years with one underlying condition (0.91%).

According to our online risk calculator, a 66-year-old man with COPD has a 6.39% (95% CI 5.48–7.30) baseline 1-year mortality risk; thus with 25 641 people in England in this subgroup, we projected 1639 baseline deaths over 1 year. The model (if set to 10% infection rate and RR 2.0) estimates 164 excess COVID-related deaths over a year in that demographic category.

At COVID-19 prevalence of 10% (mitigation) and 80% (do nothing), we estimate the number of excess deaths to be 18 374 (mitigation) and 146 996 (do nothing) with an RR of 1.5; 36 749 (mitigation) and 293 991 (do nothing) with an RR of 2.0; and 73 498 (mitigation) and 587 982 (do nothing) with an RR of 3.0. Corresponding numbers of deaths at full suppression (0.001%) and partial suppression (1%) were two (full) and 1837 (partial) with an RR of 1.5; four (full) and 3675 (partial) with an RR of 2.0; and seven (full) and 7350 (partial) with an RR of 3.0 (figure 4).

Extrapolating from our English sample to the whole UK population, we estimate 8.43 million people are in the high-risk group, of whom 5.77 million are older than 70 years of age, and 2.66 million have one or more underlying condition and are 70 years or younger). Based on the background (pre-COVID-19) 1-year mortality that we observed in the CALIBER data of 4.46% for this group, 376 001 deaths would occur in the UK. Using these estimates, numbers of people at high-risk other countries were 4.70 million in Canada, 8.31 million in France, 10.46 million in Germany, 7.51 million in Italy, 1.28 million in Sweden, and 41.76 million in the USA (table).

## Discussion

Using NHS records, we provide an initial population-based study estimating plausible ranges of excess COVID-19-related mortality in England. We provide estimates of the prevalence and background (pre-COVID-19) 1-year mortality risks according to underlying conditions at different ages in women and men, sharing these estimates in online materials, and the disease definitions in the HDR UK–CALIBER portal. These are major parameters in understanding the emergency, and the best source of estimates comes from electronic records, with implications

	General population (millions)		High-risk group (millions)			Baseline number of deaths in 1 year in high-risk group
	Total	Older than 30 years	Older than 70 years	70 years or younger with underlying condition	Total at high risk	
Canada	37.06	23.44	3.22	1.48	4.70	209 732
England	55.98	35.41	4.87	2.24	7.10	316 805
France	65.50	41.43	5.69	2.62	8.31	370 681
Germany	82.40	52.12	7.16	3.29	10.46	466 322
Italy	59.20	37.45	5.15	2.37	7.51	335 028
Sweden	10.10	6.39	0.88	0.40	1.28	57 158
UK	66.44	42.03	5.77	2.66	8.43	376 001
USA	329.10	208.17	28.60	13.16	41.76	1 862 459

All assumptions are based on England (CALIBER) estimates of 63.25% of the total population older than 30 years, 13.74% older than 70 years, 6.32% 70 years or younger with an underlying condition, and 20.06% total high-risk group; and a 1-year mortality (Kaplan-Meier) for the total high-risk group of 4.46%.

**Table: Projected deaths in 1 year in high-risk group pre-COVID-19 infection across countries**

for policy makers, clinicians, and the public. Our results support rigorous, sustained implementation of suppression measures.<sup>25</sup>

With our definition of overall excess deaths, combining direct and indirect effects among those with and without infection, it is reasonable to consider that a high (eg, 80%) proportion of the population might be affected by the COVID-19 pandemic. We do not know what the relative impact of excess deaths from the COVID-19 pandemic will be over a 1-year period in any country. However, clinical and societal concern internationally is that it is not a regular seasonal flu. We show that if the pandemic is associated with an RR of 1.5, then there will be 18 374 excess deaths in a mitigation scenario and 146 996 excess deaths in a do nothing scenario; compared with two deaths under a full suppression scenario. A doubling of the risk is worth considering because of uncertainty regarding the actual risk associated with COVID-19 and mounting health-care burden around the world. This RR of 2.0 would result in 36 749 excess deaths in a mitigation scenario and 293 991 excess deaths in a do nothing scenario.

A major concern is that relative risks will rise in a non-linear fashion with infection rates if health systems become overwhelmed by critically ill patients. Thus, at high infection rates, relative risks for excess mortality would almost certainly be much higher—and these non-linear interactions are a subject for further modelling. The impact of COVID-19 is likely to be much greater if there is poor compliance with social isolation policies and low treatment and health-system capacity. The RR of COVID-19 infection and the case fatality rate have been estimated,<sup>21</sup> but the RR for mortality will better capture the impact of infection on individuals and health systems. Even in the RR of 2.0 model, full

For our online risk calculator prototype see <http://covid19-phenomics.org/PrototypeOurRiskCoV.html>

suppression would lead to virtually no excess mortality, because some patients with underlying conditions would have died over a 1-year period regardless of the COVID-19 outbreak. The prototype online risk calculator allows exploration and visualisation of the age-specific, sex-specific, and underlying condition-specific excess mortality under different assumptions. Near real-time cause-specific mortality, as reported by the Office for National Statistics, ideally on a weekly basis, will greatly aid understanding of the direct and indirect impacts of the COVID-19 pandemic.<sup>10</sup> Such information can inform timing of exit strategies from lockdown. In future analyses, the impact of the COVID-19 pandemic on disease-specific health-care provision and use should be assessed.

We have assumed the same impact (RR) of COVID-19 on excess mortality for each underlying condition—ie, it is the same for a patient with COPD or with morbid obesity. Although a reasonable starting point, this assumption requires testing as empirical data become available. An interaction between different comorbidities almost certainly exists, meaning that some combinations are worse than others, and this factor needs to be investigated.

We used a 1-year time period because of concerns regarding second waves of infection and the time needed for health services and the wider socioeconomic disruption to return to a new steady state. Complementary future analyses might usefully address the temporal resolution (daily and weekly) of excess mortality estimates, which are likely to change at different phases of the epidemic; specifically to focus on the first 3 months, extending to 2 years and beyond.

We show that by Public Health England criteria, at least 20% of the population falls within the high-risk mortality category for COVID-19: 13·7% based on age older than 70 years (an arbitrary cutoff) and a further 6·3% based on having one or more underlying conditions. We show how policy might consider age in combination with underlying conditions. For example, a man aged 66–70 years with no underlying conditions, who is not currently considered high risk, has a higher background 1-year mortality (1·07%) than that of a woman aged 56–60 years with one underlying condition (0·91%), who is considered high risk.

The underlying conditions that we report (cardiovascular diseases, COPD, and diabetes) are known to already be suboptimally managed before COVID-19, with missed opportunities for effective (secondary) preventive interventions common in many countries.<sup>26</sup> With COVID-19 pressures, these practical actions for clinicians might not be addressed, and chronic disease management care might deteriorate. Primary and secondary health-care workers have an important role in optimising guideline-recommended management of underlying conditions to lower background risk, particularly if shielding or suppression measures are to be sustained over long periods—eg, 3 or more months. The prototype online

risk calculator might be relevant for clinicians to prioritise such patients.

In England, 1·5 million individuals have been identified based on national health records as being extremely vulnerable to COVID-19 infection, on the basis of one of a wide range of factors (including pregnancy with serious heart disease, chemotherapy, solid organ transplants, and cystic fibrosis).<sup>12</sup> These patients have been identified and are being individually supported (eg, with food parcels and medicine delivery) for the life-changing intervention of 12 weeks of shielding. Such policies, which require actionable knowledge across multiple diseases have few, if any, historical precedents; only countries with nationwide health-system data can implement such policies. The list was probably designed to identify patients who are most susceptible to infection or have a degree of immunosuppression, which cannot be quantified in a standardised way. The prognosis research strategy initiative has previously made recommendations for consistent approaches to understanding prognosis across multiple diseases and clinical specialities, the importance of which is highlighted by the COVID-19 pandemic.<sup>27</sup>

We found that patients with conditions not on the vulnerable patient list may be at as high or greater mortality than those who are. For example, cardiovascular disease is not on the vulnerable patient list and has a 1-year mortality of 6%; we found that the 1-year mortality risk of people with two or more conditions was 11%—but multiple morbidities co-occurring in the same individual do not yet qualify as extremely vulnerable. A more systematic, transparent understanding is needed of which underlying conditions are important for policy and patients. Further research is required to identify the extent to which patients with other physical conditions, frailty, and mental health conditions as well as those experiencing social exclusion (eg, homeless people and intravenous drug users) might benefit from targeted interventions.

We have deposited definitions of underlying conditions relevant to COVID-19 in the HDR UK–CALIBER open online portal. It is important for reproducible research in COVID-19 that different research groups within and between countries use clinical information from electronic health records in consistent, transparent, and validated ways to define underlying conditions and the COVID-19 syndrome (electronic health record phenotypes). It has previously been shown that in defining one disease—asthma—research groups combine clinical data in more than 60 different ways to reach a definition.<sup>28</sup> However, few efforts have been made to harmonise definitions of diseases across national electronic health records. To address the COVID-19 pandemic, international efforts are needed to harmonise definitions of underlying conditions and the clinical syndrome and progression of COVID-19 infection using routine clinical health records.



We share age-specific (5-year age bands) and sex-specific estimates of prevalence and 1-year mortality for each underlying condition. These population-based estimates might be relevant for other countries, particularly those in which population-based records are neither available nor accessible. As an illustration, we show estimates of excess deaths in different countries if the background (pre-COVID-19) mortality from England were applied. It is known, based on disease-specific registries, that mortality among people with underlying conditions differs between countries, but few studies using electronic health records compare outcomes across countries. In one study, the long-term mortality among patients who had survived a heart attack was broadly consistent across France, Sweden, the UK, and the USA,<sup>29</sup> suggesting that our findings might have international relevance.

Like other researchers, we were able to access only a sample of relevant NHS data, covering 5% of the UK population, excluding any detailed information from hospitals. Research is a key part of the COVID-19 response, and the UK Government is already acting to increase publicly accountable access to NHS health and social care data to a large number of researchers, clinicians, and policy makers nationwide. Currently researchers can access only small samples, in which linkage to even simple data on hospital admissions requires multiple approvals and can take years. It is impossible to access NHS data on the COVID-19 epidemic as a nationwide whole (it is fragmented by primary and secondary care and across countries). Despite small networks for critical care,<sup>30</sup> and coronary disease,<sup>31</sup> there are currently no national efforts to stream real-time, actionable data on COVID-19 care and (non-fatal) outcomes. With public and Government support, the NHS has an opportunity to mobilise its data to create a learning health system needed urgently for COVID-19, in which high-quality clinical information is collected. A system like this would need to include measures of disease severity, which are often absent. Beyond the COVID-19 pandemic, this system could have far-reaching, long-term benefits for patient care by accelerating knowledge generation and application.

We might have underestimated true 1-year mortality risks because physical distancing policies might themselves influence the background mortality risks—people might be less likely to access health and social care, and physical isolation has in many previous studies been shown to be associated with the onset and progression of cardiovascular and other diseases. Underlying diseases, treatment, and health-care use need to be explored in detail in the context of COVID-19. It is not known whether sudden and systematic physical distancing,<sup>32</sup> particularly if prolonged over months, has different or additional health consequences from those reported from earlier studies of isolation. Understanding how to mitigate the emergency will require accessing and

linking data about causes and consequences of the emergency across sectors, including education, economy, and transport.

We have presented our findings in a way that begins to be understandable by the general public, because of calls for information that might be relevant to an individual's risk—here defined by age, sex, and presence of an underlying condition.<sup>1</sup> No authoritative public-facing tools are yet available, but they might provide context, motivation, and support to comply with the multiple changes in daily living that shielding entails. Particularly over prolonged isolation, this could be important given concern about the practical ability of patients to maintain shielding.

By contrast with the daily news feed in every country, which reports the numbers of deaths (the numerator), we report a risk: numbers of deaths over 1 year divided by the population who was at risk of dying. Among the public it might not be widely appreciated that 1400 people die on average every day in the UK at baseline (ie, before COVID-19).<sup>33</sup> Of all those people dying within 1 year, it is likely that COVID-19 brings forward the death earlier in the year. In other words, there are competing causes for the mortality.

Our study has important limitations. The study was developed over a 72-h period before posting the results on March 22, 2020, the day before the UK Prime Minister announced lockdown. First, our models require further data for their development. Commonly, research using NHS data can take months, and even years, partly because of the multiple steps in accessing such data for research. Specifically, it is important to study a wider range of underlying conditions, a wider range of hospital and critical care admissions, linkage of primary care data to nationally available information on COVID-19 testing, linkage to richer, real-time hospital data, and use of nationwide NHS data. Second, we believe it is both a limitation and a strength that our model is simple and supports public understanding. The assumptions we make can be tested as empirical data become available, such as the Office for National Statistics data.<sup>10</sup> There are many avenues for further modelling to better understand how to target different preventive interventions. An (incomplete) list includes statistical (dynamic models, weekly rates of mortality, and competing risks), public health (disease-specific data for high-risk and extremely vulnerable conditions, regional variation, social deprivation, and ethnicity), clinical (including hospital and critical care admissions), and health service factors (including data on health-care workers and operational features).

Health-care professionals, like those involved in this paper, are being redirected to provide priority clinical services during the emergency internationally. We believe an important aspect of emergency response is to support clinical academics to continue to contribute to research into new preventive and treatment options, as they take on new service responsibilities.

If the relative risk of COVID-19 on mortality over 1 year were 1·2 (ie, the same magnitude as the winter vs summer mortality excess) then the number of excess deaths would be zero when one in 100 000 individuals are infected (full suppression scenario), 7350 when one in ten are infected (mitigate scenario), and 58798 when eight in ten are infected (do nothing scenario). However, if the relative risk is doubled to 2·0, then there would be 36749 deaths in the mitigation scenario and 293991 deaths in the do nothing scenario. These results should inform targeting of those at highest risk for a range of preventive interventions.

#### Contributors

AB conceived the research question. SD and HH obtained funding. AB, LP, CL, SD, and HH designed the study and analysis plan. SD, AG-I, AT, WKW, and NS prepared the data, including electronic health record phenotyping in the CALIBER open portal. LP and AB did the statistical analysis. AB and HH drafted the initial and final versions of the manuscript. All authors critically reviewed early and final versions of the manuscript.

#### Declaration of interests

AB is supported by research funding from the National Institute for Health Research (NIHR), British Medical Association, AstraZeneca, and UK Research and Innovation. All other authors declare no competing interests.

#### Data sharing

We will make available the estimates of age-specific, sex-specific, and condition-specific prevalence and 1 year mortality estimates online, the electronic health record phenotypes via the HDR UK–CALIBER online portal, and the online risk calculator.

#### Acknowledgments

BW and HH are funded by the National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre. HH is supported by Health Data Research UK (HDR-UK; grant no LOND1), which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation, and Wellcome Trust. HH is also supported by the BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 joint undertaking under grant agreement no 116074. This joint undertaking receives support from the EU's Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations. This work was supported by an NIHR Clinician Scientist award (CS-2016-007) to LS. NS is supported by NIHR Great Ormond Street Hospital Biomedical Research Centre and HDR-UK.

#### References

- Spiegelhalter D. How much 'normal' risk does Covid represent? Medium, March 21, 2020. <https://medium.com/wintoncentre/how-much-normal-risk-does-covid-represent-4539118e1196> (accessed April 16, 2020).
- Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. London: Imperial College, March 16, 2020. <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf> (accessed April 16, 2020).
- WHO. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). Geneva: World Health Organization, Feb 16–24, 2020. <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf> (accessed April 16, 2020).
- Yang J, Zheng Y, Gou X, et al. Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. *Int J Infect Dis* 2020; published online March 12. DOI:10.1016/j.ijid.2020.03.017.
- WHO Emergency Committee. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Geneva: World Health Organization, Jan 30, 2020. [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)) (accessed April 16, 2020).
- Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* 2020; published online March 23. DOI:10.1001/jama.2020.4683.
- Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395: 497–506.
- Bialek S, Boundy E, Bowen V, et al. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. *MMWR Morb Mortal Wkly Rep* 2020; 69: 343–46.
- Oke J, Heneghan C. Global COVID-19 case fatality rates. CEBM Research, April 6, 2020. <https://www.cebm.net/global-covid-19-case-fatality-rates/> (accessed April 16, 2020).
- Office for National Statistics. Deaths registered weekly in England and Wales, provisional: week ending 3 April 2020. April 14, 2020. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending3april2020#main-points> (accessed April 17, 2020).
- US Centers for Disease Control and Prevention. Coronavirus disease 2019 (COVID-19). People who are at higher risk. April 15, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/specific-groups/high-risk-complications.html> (accessed April 16, 2020).
- Public Health England. Guidance on social distancing for everyone in the UK. March 30, 2020. <https://www.gov.uk/government/publications/covid-19-guidance-on-social-distancing-and-for-vulnerable-people/guidance-on-social-distancing-for-everyone-in-the-uk-and-protecting-older-people-and-vulnerable-adults> (accessed April 16, 2020).
- UK Government. Major new measures to protect people at highest risk from coronavirus. March 21, 2020. <https://www.gov.uk/government/news/major-new-measures-to-protect-people-at-highest-risk-from-coronavirus> (accessed April 16, 2020).
- Denaxas S, Gonzalez-Izquierdo A, Direk K, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc* 2019; 26: 1545–59.
- Kuan V, Denaxas S, Gonzalez-Izquierdo A, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digital Health* 2019; 1: e63–77.
- Rapsomaniki E, Timmis A, George J, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. *Lancet* 2014; 383: 1899–911.
- Pikoula M, Quint JK, Nissen F, Hemingway H, Smeeth L, Denaxas S. Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. *BMC Med Inform Decis Mak* 2019; 19: 86.
- Hurst JR, Dickhaus J, Maulik PK, et al. Global Alliance for Chronic Disease researchers' statement on multimorbidity. *Lancet Glob Health* 2018; 6: e1270–71.
- Bozkurt B, Kovacs R, Harrington R. HFSA/ACC/AHA statement addresses concerns re: using RAAS antagonists in COVID-19. March 17, 2020. [https://professional.heart.org/professional/ScienceNews/UCM\\_505836\\_HFSAACCAHAsstatement-addresses-concerns-re-using-RAAS-antagonists-in-COVID.jsp](https://professional.heart.org/professional/ScienceNews/UCM_505836_HFSAACCAHAsstatement-addresses-concerns-re-using-RAAS-antagonists-in-COVID.jsp) (accessed April 16, 2020).
- US Food and Drug Administration. FDA advises patients on use of non-steroidal anti-inflammatory drugs (NSAIDs) for COVID-19. March 19, 2020. <https://www.fda.gov/drugs/drug-safety-and-availability/fda-advises-patients-use-non-steroidal-anti-inflammatory-drugs-nsaids-covid-19> (accessed April 16, 2020).
- Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis* 2020; published online March 12. [https://doi.org/10.1016/S1473-3099\(20\)30195-X](https://doi.org/10.1016/S1473-3099(20)30195-X).

- 22 Lintern S. Coronavirus could kill half a million Britons and infect 80% of UK population, government documents indicate. *The Independent*, Feb 26, 2020. <https://www.independent.co.uk/news/health/coronavirus-news-latest-deaths-uk-infection-flu-a9360271.html> (accessed April 16, 2020).
- 23 Office for National Statistics. Deaths by single year of age tables, UK. Jan 17, 2020. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathregistrationssummarytablesenglandandwalesdeathsbyingleyearofagetables> (accessed April 16, 2020).
- 24 United Nations Population Fund. World population dashboard. <https://www.unfpa.org/data/world-population-dashboard> (accessed April 16, 2020).
- 25 Cyranoski D. What China's coronavirus response can teach the rest of the world. *Nature* 2020; **579**: 479–80.
- 26 Yusuf S, Islam S, Chow CK, et al. Use of secondary prevention drugs for cardiovascular disease in the community in high-income, middle-income, and low-income countries (the PURE study): a prospective epidemiological survey. *Lancet* 2011; **378**: 1231–43.
- 27 Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013; **346**: e5595.
- 28 Al Sallakh MA, Vasileiou E, Rodgers SE, Lyons RA, Sheikh A, Davies GA. Defining asthma and assessing asthma outcomes using electronic health record data: a systematic scoping review. *Eur Respir J* 2017; **49**: 1700204.
- 29 Rapsomaniki E, Thuresson M, Yang E, et al. Using big data from health records from four countries to evaluate chronic disease outcomes: a study in 114 364 survivors of myocardial infarction. *Eur Heart J Qual Care Clin Outcomes* 2016; **2**: 172–83.
- 30 Harris S, Shi S, Brealey D, et al. Critical Care Health Informatics Collaborative (CCHIC): data, tools and methods for reproducible research: a multi-centre UK intensive care database. *Int J Med Inform* 2018; **112**: 82–89.
- 31 Kaura A, Panoulas V, Glampson B, et al. Association of troponin level and age with mortality in 250 000 patients: cohort study across five UK acute care centres. *BMJ* 2019; **367**: l6055.
- 32 Fenichel EP. Economic considerations for social distancing and behavioral based policies during an epidemic. *J Health Econ* 2013; **32**: 440–51.
- 33 BBC News. Coronavirus: UK deaths double in 24 hours. March 14, 2020. <https://www.bbc.co.uk/news/uk-51889957> (accessed April 16, 2020).