BMJ Open

COVID-19 mortality effects of underlying health conditions in India: a modelling study

Paul Novosad [1], Radhika Jain [2], Alison Campion [3], Sam Asher [4]

To cite: Novosad P, Jain R, Campion A, *et al*. COVID-19 mortality effects of underlying health conditions in India: a modelling study. *BMJ Open* 2020;**10**:e043165. doi:10.1136/bmjopen-2020-043165

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-043165).

Received 29 July 2020
Revised 02 November 2020
Accepted 25 November 2020

[1]Economics, Dartmouth College, Hanover, New Hampshire, USA
[2]Stanford University, Stanford, California, USA
[3]Development Data Lab, Washington, DC, USA
[4]International Economics, Johns Hopkins University Paul H Nitze School of Advanced International Studies, Washington, DC, USA

Correspondence to
Dr Paul Novosad;
paul.novosad@dartmouth.edu

ABSTRACT

Objective To model how known COVID-19 comorbidities affect mortality rates and the age distribution of mortality in a large lower-middle-income country (India), and to identify which health conditions drive differences with high-income countries.

Design Modelling study.

Setting England and India.

Participants Individual data were obtained from the fourth round of the District Level Household Survey and Annual Health Survey in India, and aggregate data were obtained from the Health Survey for England and the Global Burden of Disease, Risk Factors and Injuries Studies.

Main outcome measures The primary outcome was the modelled age-specific mortality in each country due to each COVID-19 mortality risk factor (diabetes, hypertension, obesity and respiratory illness, among others). The change in overall mortality and in the share of deaths under age 60 from the combination of risk factors was estimated in each country.

Results Relative to England, Indians have higher rates of diabetes (10.6% vs 8.5%) and chronic respiratory disease (4.8% vs 2.5%), and lower rates of obesity (4.4% vs 27.9%), chronic heart disease (4.4% vs 5.9%) and cancer (0.3% vs 2.8%). Population COVID-19 mortality in India, relative to England, is most increased by uncontrolled diabetes (+5.67%) and chronic respiratory disease (+1.88%), and most reduced by obesity (–5.47%), cancer (–3.65%) and chronic heart disease (–1.20%). Comorbidities were associated with a 6.26% lower risk of mortality in India compared with England. Demographics and population health explain a third of the difference in share of deaths under age 60 between the two countries.

Conclusions Known COVID-19 health risk factors are not expected to have a large effect on mortality or its age distribution in India relative to England. The high share of COVID-19 deaths from people under age 60 in low- and middle-income countries (LMICs) remains unexplained. Understanding the mortality risk associated with health conditions prevalent in LMICs, such as malnutrition and HIV/AIDS, is essential for understanding differential mortality.

Strengths and limitations of this study

- This study modelled the extent to which demographics and the population prevalence of COVID-19 mortality risk factors are likely to affect mortality in India, a major lower-middle-income country.
- This study distinguished between controlled and uncontrolled health conditions (such as diabetes), a difference that may significantly affect population risk.
- This study used multivariate HRs from the largest sample analysis of risk factors for COVID-19 mortality to date (N=17 278 392), which is essential because bivariate HRs may overestimate the effect of pre-existing health conditions that increase with age.
- In addition to modelling total mortality, this study modelled the age distribution of deaths, which has been very different in low- and middle-income countries (LMICs) and high-income countries.
- The key limitation of this study is that there are almost no large sample analyses describing the additional risk of COVID-19 mortality associated with health conditions that are more common in LMICs than in high-income countries, such as protein calorie malnutrition, micronutrient deficiency and HIV/AIDS.

INTRODUCTION

The number of cases of COVID-19 continues to rise around the world. A growing share of cases is coming from low- and middle-income countries (LMICs) in Asia, Africa and the Americas that were largely spared in the initial stages of the pandemic.[1] Because the severity of infection increases substantially with age, forecasts have projected much lower aggregate mortality rates in LMICs than in high-income countries.[2–4]

However, the reported fatality numbers from LMICs to date have suggested a much greater share of COVID-19 deaths among the young. As of May 2020, 30.5% of deaths in Brazil occurred in those under age 60, whereas 27% of deaths in Mexico occurred in those under age 50.[5–7] As of July 2020, 47% of COVID-19 deaths in India occurred in those under age 60, whereas a particular study of Tamil Nadu and Andhra Pradesh found 54% of deaths occurred in those under age 65.[8 9]

In contrast, individuals under age 65 have accounted for only 5%–13% of deaths in 10 European countries and Canada and 8%–24% in US states.[10] It is not presently known whether the different age pattern of deaths in LMICs is driven by erroneous reporting, differences in infection patterns, younger populations or different rates of COVID-19 comorbidities in the underlying populations.

Many modelling studies have presumed that worse population health in LMICs will lead to excess mortality or else have ignored differential population health as a factor entirely.[4 11 12] To date, there has been limited analysis of the prevalence in LMICs of the specific conditions associated with increased COVID-19 severity, such as diabetes, obesity, cardiovascular disease, hypertension and chronic kidney disease, or of how they change the expected level and age distribution of mortality.[4 13–16] Some studies have adjusted mortality estimates for population comorbidities by treating all comorbidities as equivalent or by multiplying the mortality rate by a fixed amount to adjust for population health.[17–20] One study combined condition-specific prevalence and HRs from a sample of hospitalisations, but excluded obesity and uncontrolled diabetes, and did not examine mortality or the age distribution of mortality as outcomes.[21]

Using England as a benchmark, this study examines how comorbidities understood to increase COVID-19 mortality are likely to affect COVID-19 mortality rates in aggregate and across the age distribution in India, identifying the specific risk factors with the largest mortality effects. We further study the extent to which accounting for differences in demographics and underlying health conditions can explain the increased share of deaths among the young in India relative to England.

Our analysis focuses on India and on the COVID-19 risk factors that are currently documented. At the time of writing, India has the second highest number of cumulative COVID-19 infections in the world and one of the highest growth rates in infections of any major country, making it an essential population to study.[1] The methodology is readily adjusted to account for new risk factors or data from other countries and may be useful for modelling the epidemic in a range of LMICs.

METHODS

Our approach requires three types of data: (1) the relative risk of COVID-19 mortality associated with gender, age and each health condition; (2) the age-specific prevalence of each health condition in England and India; and (3) the age and gender distributions for the two countries.

Estimates of relative risk of COVID-19 mortality from comorbidities

We obtained estimates of COVID-19 mortality risk for a wide range of comorbidities from the OpenSAFELY study, a closed cohort study of 17 278 392 adults from England.[22] This was the largest analysis of comorbidities associated with COVID-19 mortality to date and one of the few studies that estimates risk factors in a multivariate model adjusting for age, sex and other health conditions. This adjustment is important because many COVID-19 comorbidities are increasing in age and their HRs are thus biased upwards in analyses not adjusting for age.

The OpenSAFELY study enrolled all individuals registered with a general practice within The Phoenix Partnership system on 1 February 2020, who were aged 18 years or older on enrolment, who had at least 1 year of medical history within the system and who had recorded age and sex. The underlying data set represents 40% of the population of England, and the prevalence of health conditions in the study cohort is similar to estimates of population prevalence in England (online supplemental appendix p 4). Patients were followed through 6 May. The outcome was in-hospital death among people with confirmed COVID-19 infections. HRs for mortality from a Cox proportional-hazards model were estimated for a comprehensive list of risk factors described in other studies, adjusted for sex, age and all other risk factors. As patient-level data from OpenSAFELY are not publicly available, we extracted HRs from the paper reporting results of the analysis.[22] In the absence of patient-level COVID-19 mortality data with comorbidity information in India, we assumed that comorbidity-associated HRs were the same in India as in England.

Ideally, HRs would measure mortality risk conditional on infection, rather than on registration with a general practice (as in OpenSAFELY) or hospitalisation (as in prior work).[21 23] The HRs in this study therefore reflect combined mortality and infection risk; the analysis assumes that pre-existing health conditions are not significant predictors of infection. Reassuringly, HRs measuring mortality risk conditional on hospitalisation in other studies are similar to those used here.[11]

Demographics and risk factor prevalence in India and England

Age distributions and age-specific sex ratios for India and England were obtained from official censuses.

We obtained data on age-specific prevalence of health risk factors for India and England from multiple sources, prioritising biomarker data where available and matching definitions as closely as possible to the conditions for which HRs are available. We restrict samples to ages 18–99 for consistency with the HRs.

For India, we aggregated individual-level biomarker data from two public population health surveys for obesity, diabetes and hypertension. The fourth round of the Indian District Level Household Survey and the second round of the Annual Health Survey were conducted between 2012 and 2014; they jointly cover 94% of the Indian population and provide the most recent nationwide direct measures of height, weight, fasting plasma glucose (FPG) and blood pressure (BP) for adults of all ages in India. Details of data set construction are provided in online supplemental appendix p 1.

For England, age-specific prevalences of obesity, hypertension and diabetes were obtained from the nationally representative 2018 Health Survey for England, which collected symptoms and medical diagnoses for a range of illnesses, as well as direct measures of height, weight, BP and glycated haemoglobin (HbA1c).[24]

Body mass index was classified into no evidence of obesity (<30 kg/m^2), obese class 1 or 2 (30–39.9 kg/m^2) and obese class 3 (40+ kg/m^2). Hypertension was defined as systolic BP ≥140 mm Hg or diastolic BP ≥90 mm Hg (uncontrolled) or a medical diagnosis of hypertension with BP below the thresholds (controlled). The prevalence of controlled and uncontrolled hypertension was reported separately but combined in the risk estimation for consistency with OpenSAFELY. OpenSAFELY classified controlled diabetes as HbA1c >51 mmol/mol and HbA1c <58 mmol/mol, and uncontrolled diabetes as HbA1c ≥58 mmol/mol. Corresponding thresholds for the one-time FPG measures in the Indian data set are not well defined. In England, the prevalence was reported based on a threshold of HbA1c ≥48 mmol/mol (6.5%). Therefore, we followed the standard screening and diagnosis thresholds recommended by the WHO and International Diabetes Federation and defined uncontrolled diabetes in India as a plasma glucose reading ≥126 mg/dL (7.0 mmol/L) if fasting or ≥200 mg/dL (11.1 mmol/L) if not fasting. We used the corresponding recommended threshold of HbA1c ≥48 mmol/mol (6.5%) for uncontrolled diabetes in England.[25] In both countries, we classified individuals with biomarkers below the thresholds but with a diagnosis of diabetes as having controlled diabetes.

Age-specific prevalence for asthma, chronic heart disease, kidney disease, stroke, dementia, haematological malignancies and all other cancers was drawn from the Global Burden of Diseases, Risk Factors and Injuries Studies (GBD) for India and England.[26] OpenSAFELY reports separate HRs for cancers diagnosed <1 year ago, 1–4.9 years ago and ≥5 years ago; because the year of diagnosis is unavailable in GBD, we used a single classification for each class of cancers and the HR for diagnosis <1 year ago. For chronic respiratory disease, we used chronic obstructive pulmonary disease (COPD) prevalence from the GBD for India and modelled COPD prevalence from the Clinical Practice Research Datalink cohort database for England.[27] GBD prevalence of Parkinson's disease, epilepsy, multiple sclerosis and motor neuron disease was combined and classified as neurological disorders.

The following risk factors were not available for India and were excluded from the analysis for both England and India for comparability: fibrosing lung disease, bronchiectasis or cystic fibrosis, lupus, asthma with no recent oral corticosteroid use, cancers diagnosed more than a year ago, organ transplant and spleen disease. Given that the relationship between smoking and COVID-19 mortality remains under debate, we excluded it from the analysis.[28] We also excluded ethnicity and socioeconomic status, which cannot be measured comparably across England and India and are unlikely to have similar relative risk in the two countries. Where aggregate data were available only in coarse age bins, we used non-linear interpolation to reaggregate the data into age bins matching the HR data.

Estimating the contribution of health conditions to population COVID-19 mortality risk

The OpenSAFELY study reports HRs for each age group, sex and health condition with women aged 50–59 years with no conditions as the reference group.[22] We transform the HR for each health condition c into a relative risk (RR_c) assuming a population mortality rate r of 1%:

$$RR_c = \left(1 - e^{HR_c \ln(1-r)}\right) / r$$

To obtain continuous relative risk for age, we used a polynomial interpolation for the log HR at each age, renormalising with age 50 as the reference group (online supplemental appendix p 5).

The additional population mortality risk associated with a given health condition increases with the condition's relative risk for COVID-19 mortality and with its prevalence. We defined the age-specific and condition-specific population relative risk $\mathrm{PRR}_{a,c}$ of condition c at age a as follows:

$$PRR_{a,c} = RR_c \times PREV_{a,c} + (1 - PREV_{a,c})$$

where $PREV_{a,c}$ is the prevalence of condition c at age a, and $PRR_{a,c}$ describes the proportional increase in mortality at age a associated with health condition c.

We combined PRRs to obtain an age-specific population relative risk of mortality arising from the combined prevalence of all the health conditions:

$$PRR_a = \prod_{c \in C} PRR_{a,c}$$

PRR_a isolates the expected mortality difference at each age between India and England that is driven by the combined prevalence of all the health conditions studied. This approach implicitly assumes that the health conditions are uncorrelated with each other. Without microdata on the full set of health conditions, this assumption is unavoidable, but will bias the England versus India comparison only if the correlation of health conditions is substantially different in the two countries. We explore the possible extent of this bias in the online supplemental appendix. Using age-specific prevalence, our analysis fully accounts for the substantial correlations between age and health conditions.

We next calculated the increase in population mortality from each health condition across all ages, taking into account the age-specific prevalence of each health condition, its relative risk and the population share at each age. The condition-specific population relative risk of each health condition across the full population (PRR_c) is given by the following equation:

$$PRR_c = \sum_{a \epsilon [18,99]} \left(PRR_{a,c} * POPSHARE_a\right)$$

PRR_c is greater when the relative risk of condition c is higher (RR_c) and when its prevalence is higher at ages with higher population. The combined effect on population mortality of all the health conditions is given by the product of each condition-specific PRR_c.

Finally, we aggregated the population relative risks across health conditions to model the age distribution of deaths in each country. The number of deaths at each age N_a is the product of the mortality rate of the reference group (50-year-old women with no other risk factors), the population at age a, the age-specific population relative risk of the full set of health conditions (PRR_a), the PRR of gender and the direct relative risk of COVID-19 mortality for an individual at age a (RR_a):

$$N_a = r \times POP_a \times PRR_a \times PRR_{a,\ male} \times RR_a$$

We plotted the age distribution of deaths as shares of all deaths rather than in levels, eliminating the need to assume a reference group mortality rate. We summarised the shape of the distribution by reporting the share of expected deaths in each country that are under the age of 60. We present results from three models: (1) England's demographics and health distribution, (2) India's demographics and health distribution and (3) India's demographics but England's age-specific prevalence of health risk factors. The third model allowed us to examine the mortality shift that comes from differences in population health alone.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation or writing. The corresponding author had full access to the data in the study and had final responsibility for the decision to submit for publication.

Patient and public involvement

Because this study uses existing epidemiological data, it was not appropriate to involve patients or the public in the research.

RESULTS

Prevalence of COVID-19 risk factors in India and England

Demographic characteristics and overall prevalence of risk factors are substantially different in India relative to England (table 1). In all, 83.7% of Indian adults are below the age of 60, compared with 69.9% of English adults. Indians have substantially lower rates of obesity and cancer (4.4% and 0.3% in India compared with 27.9% and 2.8% in England), but higher rates of uncontrolled diabetes, kidney disease and chronic liver disease (8.9%, 9.7% and 5.3% in India compared with 2.1%, 5.6% and 2.6% in England).

We show differences in age-specific prevalence between India and England for the conditions for which we have biomarkers in India and are more precisely estimated (figure 1), as well as age-specific prevalence of all conditions for both countries (online supplemental appendix p 3). The overall rates of diabetes are higher in India at all ages, but diabetes in India is overwhelmingly uncontrolled, whereas three-quarters of diabetes is controlled in England. Hypertension (the sum of controlled and uncontrolled) is higher in India at young ages (31.3% for ages 40–49 in India and 18.3% in England) but lower at higher ages (52.3% at ages 70–79 in India and 61.3% in England) and is overwhelmingly uncontrolled. Conversely, obesity rates are higher at all ages in England.

Table 1 Prevalence of COVID-19 risk factors in India and England

	Prevalence (%)	
	India	England
Ages 18–39	50.2	36.6
Ages 40–49	19.2	16.3
Ages 50–59	14.3	17.0
Ages 60–69	10.3	13.3
Ages 70–79	4.6	10.4
Ages 80–99	1.5	6.3
Male	47.1	48.9
Diabetes (controlled)	1.7	6.4
Diabetes (uncontrolled)	8.9	2.1
Hypertension	28.2	28.0
Obese (classes 1 and 2)	4.0	24.8
Obese (class 3)	0.4	3.1
Chronic heart disease	4.4	5.9
Chronic respiratory disease	4.8	2.5
Asthma	2.5	9.2
Kidney disease	9.7	5.6
Chronic liver disease	5.3	2.6
Haematological cancer	0.0	0.2
Non-haematological cancer	0.3	2.6
Stroke, dementia	1.3	1.5
Other neurological conditions	0.0	0.1
Psoriasis, rheumatoid	1.0	2.4
Other immunosuppressive conditions	0.1	0.1

Relative risk of COVID-19 mortality from combined risk factors in India and England

The age-specific population relative risk of COVID-19 mortality from all health conditions combined (PRR_a) is higher in India than in England at nearly all ages, but the difference in PRR_a between the two countries is below 15% at every age and highest between ages 40 and 80 (figure 2).

PRR_a reflects the age-specific prevalence and associated COVID-19 mortality risk from all health conditions combined. Taking risk, prevalence and population at every age into account provides the full population relative risk of COVID-19 mortality from each health condition (PRR_c)—or the proportional increase in population

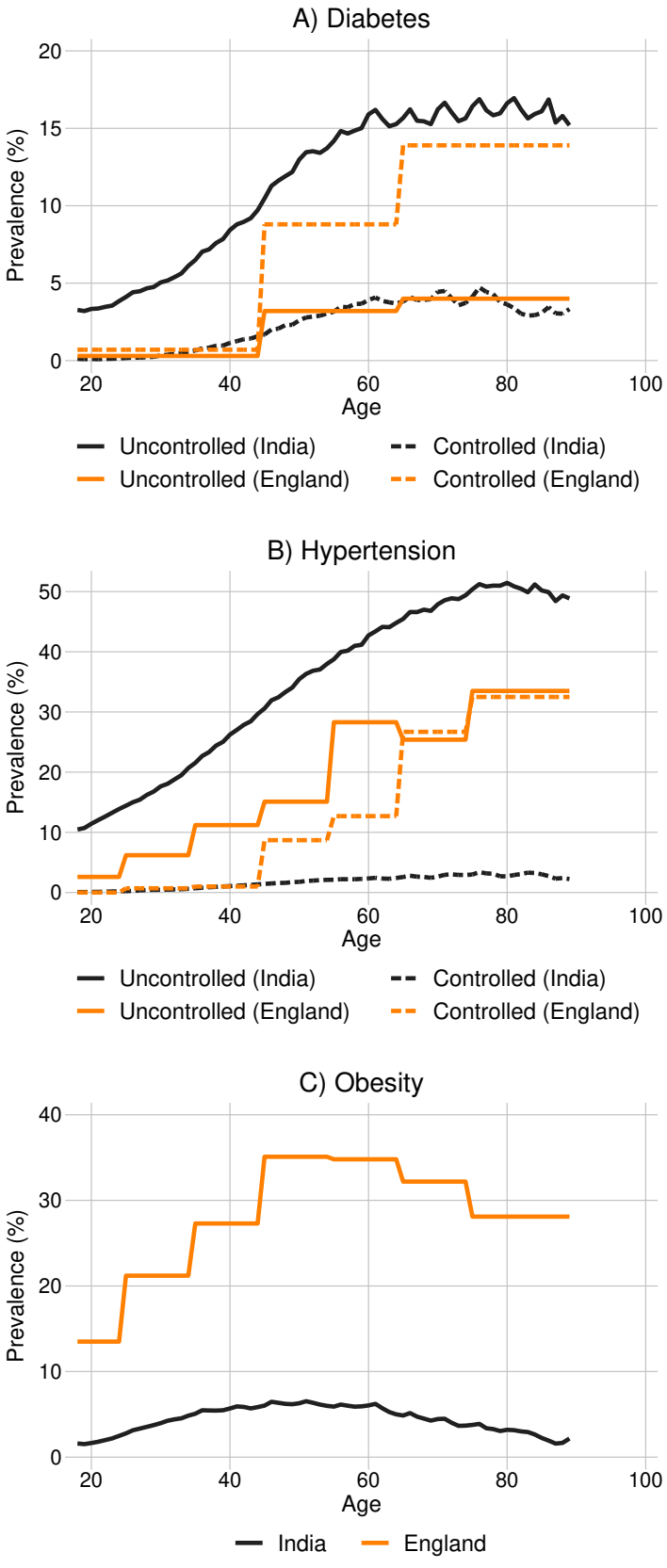


Figure 1 Prevalence of (A) diabetes, (B) hypertension and (C) obesity in India and England.

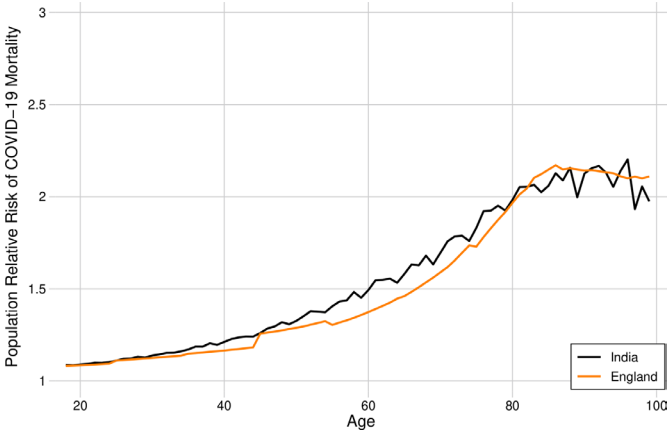


Figure 2 Age-specific population relative risk of COVID-19 mortality from all health conditions (PRR_a).

mortality across all ages associated with each health condition (table 2).

Uncontrolled diabetes, which is associated with substantial mortality risk (RR_c=1.94), increases total mortality by 7.8% in India (PRR_c=1.078), but only by 2.0% in England (PRR_c =1.020), reflecting its significantly higher prevalence in India at all ages. In contrast, controlled diabetes, more common in England than in India, is associated with 2.0% higher mortality in England and only 0.4% in India (PRR_c=1.020 vs 1.004). In addition to uncontrolled diabetes, the health conditions associated with the largest increases in mortality in India are kidney disease (PRR_c =1.050) and chronic respiratory disease (PRR_c=1.035). In England, the most consequential health conditions are obesity (combined PRR_c across all obesity classes=1.065) and kidney disease (PRR_c=1.046).

Table 2 Population relative risk of COVID-19 mortality from each health condition (*PRRc*) .

	Individual relative risk	Population relative risk (PRR_c)	
		India	England
Diabetes (controlled)	1.31	1.004	1.020
Diabetes (uncontrolled)	1.94	1.078	1.020
Hypertension	0.89	0.971	0.969
Obese (classes 1 and 2)	1.15	1.006	1.037
Obese (class 3)	1.91	1.004	1.028
Chronic heart disease	1.17	1.008	1.021
Chronic respiratory disease	1.62	1.035	1.015
Asthma	1.13	1.003	1.011
Kidney disease	1.42	1.050	1.046
Chronic liver disease	1.73	1.042	1.024
Haematological cancer	2.79	1.000	1.007
Non-haematological cancer	1.71	1.002	1.033
Stroke, dementia	2.15	1.016	1.041
Other neurological conditions	2.56	1.002	1.002
Psoriasis, rheumatoid	1.19	1.002	1.007
Other immunosuppressive conditions	1.69	1.001	1.001

Comparing the percentage difference between PRR_c of each health condition between India and England (figure 3), the condition with the largest differential impact on mortality between the two countries is uncontrolled diabetes, which is associated with 5.67% higher population mortality in India relative to England. Mortality in India relative to England is also increased by chronic respiratory disease (+1.88%) and chronic liver disease (+1.73%), but decreased by the differential prevalence of obesity (combined −5.47%), cancer (−3.65%) and stroke/dementia (−2.47%). No other risk factor has an effect of greater than ±2% on India's relative mortality. The combined effect of health conditions leads to 6.26% *higher* mortality in England than in India, reflecting England's higher age-specific prevalence of certain conditions like obesity and cancer, as well as its older age structure that increases population share at ages with worse health. This differential mortality risk does *not* include the direct effect of older age, which is associated with substantial risk (RR_a=6.08 for ages 70–80 and 20.61 for age >80) and magnifies England's mortality disadvantage substantially.

Combining the population relative risk from health conditions with the direct effect of demographics on mortality, we modelled the density of deaths across the age distribution (figure 4). In England, 9.4% of modelled deaths are below age 60, closely matching the 6.5% observed in England through May 2020 and reported in the OpenSAFELY data set. In India, 23.5% of modelled deaths are below age 60, which is substantially lower than the 50% observed in case reports. Applying England's age-specific prevalence of health conditions to India's demographic distribution, to isolate the effect of health conditions from demographics, results in a distribution nearly identical to the India model. In other words, differences in health conditions between India and England have almost no effect on mortality, indicating that the higher share of deaths in younger populations in India comes from the demographic distribution alone.

In the online supplemental appendix, we test sensitivity to uncertainty in prevalences and HRs (online supplemental appendix pp 6–8). The latter estimates cover alternate HRs estimated from other studies.[21 29] We also test sensitivity to alternate assumptions about covariance of health conditions (online supplemental appendix p 9). In all cases, we find that the population relative risk from health conditions in England is greater than that in India, and that accounting for health conditions cannot explain any of the higher incidence of mortality among the young in India relative to England.

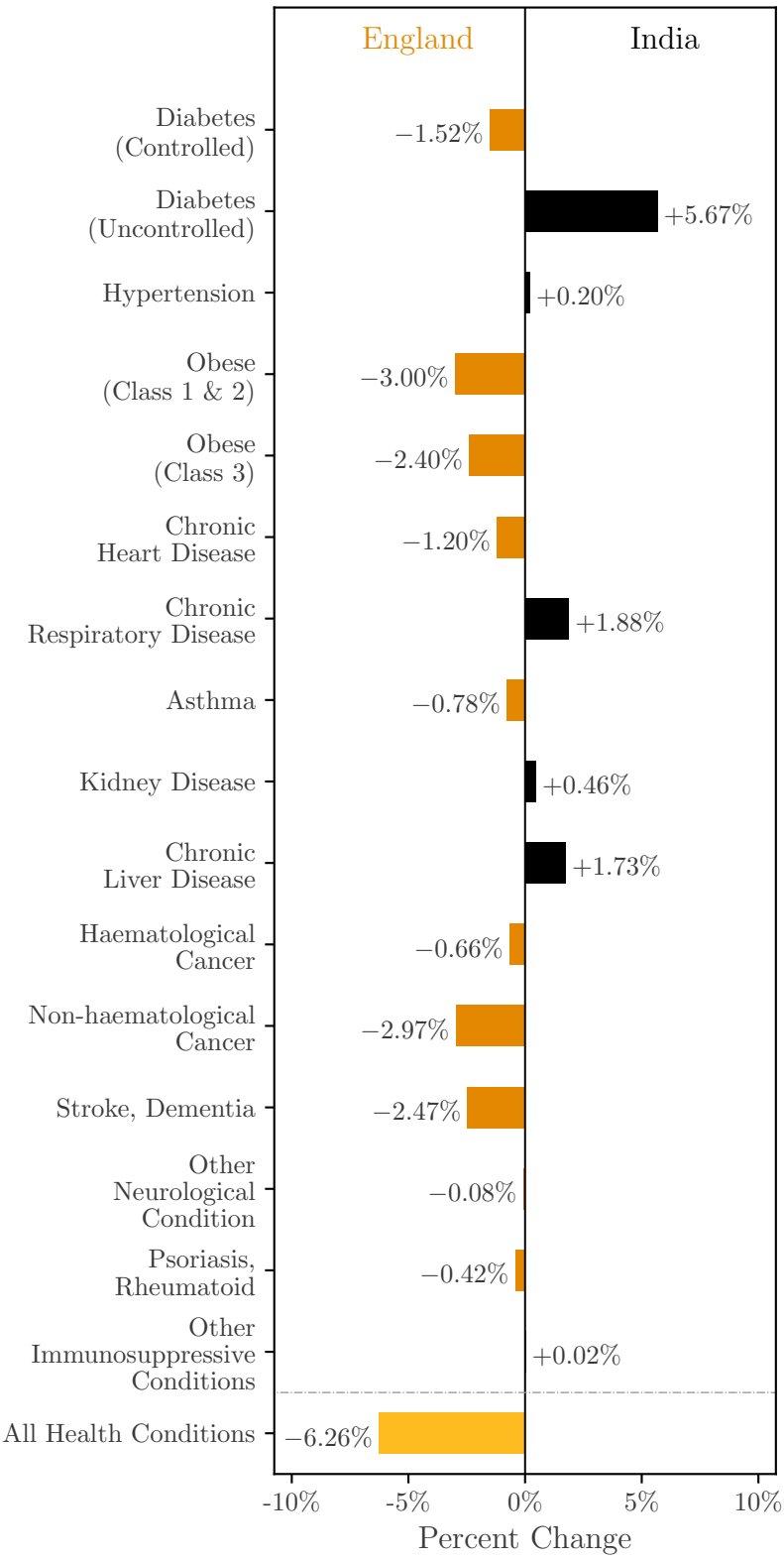


Figure 3 Percent change in population relative risk of COVID-19 mortality from each health condition (PRR_c) in India versus England.

DISCUSSION

We used the best publicly available data on population health to examine the extent to which demographics and pre-existing health conditions known to be associated with COVID-19 mortality can account for the disproportionately high share of COVID-19 deaths in younger populations observed in India relative to England. We show that differences in population health do not

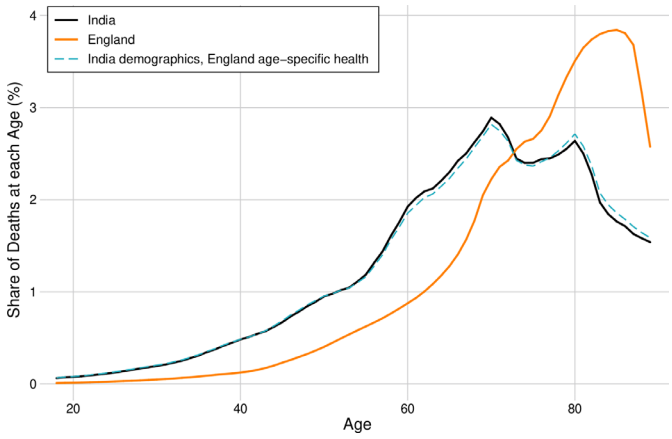


Figure 4 Modelled age distribution of COVID-19 mortality.

significantly shift the relative age distribution of disease severity and slightly *lower* aggregate mortality in India relative to England. Higher prevalence of diabetes and respiratory illness increases mortality risk relative to England, but these effects are offset by lower rates of obesity, heart disease and cancer. While the Indian age distribution substantially shifts expected mortality towards the young, it explains only a third of the difference in the share of deaths under 60 compared with England.

Epidemiological models have typically assumed that comorbidities will exacerbate the mortality of COVID-19 in India and other LMICs relative to high-income countries. We found that comorbidities identified as key risk factors in high-income countries are not associated with higher expected mortality in India relative to England, in aggregate or among the young. This suggests that understanding the other factors that may explain the differential mortality among the young observed in lower income contexts, such as different patterns of infection, under-resourced health systems or comorbidities unique to LMICs, should be a priority for further research.

This study improves on prior work by examining the extent to which comorbidities can explain the younger incidence of COVID-19 mortality in LMICs, by estimating mortality effects of specific comorbidities and by calibrating a model with a comprehensive set of comorbidity HRs drawn from a large-sample multivariate analysis of COVID-19 mortality. Models calibrated with bivariate HRs or raw prevalences of comorbidities among severe cases are likely to overestimate the effect of pre-existing health conditions because of the significant increase in all comorbidities with age alongside the direct effect of age on COVID-19 mortality.

The key limitation of this study is that there are virtually no data on the COVID-19 mortality risks associated with health conditions that are more common in LMICs than in high-income countries, such as protein calorie malnutrition, micronutrient deficiency and HIV/AIDS.[4] If these conditions make individuals more susceptible to severe infections, then population health may indeed exacerbate the severity of COVID-19 in LMICs. Understanding the extent to which health conditions endemic to LMICs affect COVID-19 severity is an urgent priority, particularly as policy responses increasingly focus on identifying and isolating high-risk individuals.[30]

Our analysis is also constrained by the limited and changing evidence on risk factors for COVID-19 severity. Based on the availability of existing measures, our model assumed that health condition relative risks are age-invariant. However, data from New York's epidemiological surveillance system suggest that hypertension and diabetes may contribute more to mortality at younger ages,[31] which would exacerbate the burden of illness among the young in LMICs. Furthermore, if illness severity and the quality of prior medical management of pre-existing health conditions change mortality risk for the same diagnosis across contexts, applying HRs from England may understate mortality risk in India. Finally, HRs which are not conditioned on infection may reflect infection risk in addition to disease severity risk and thus may not translate directly to the Indian context.

Recognising these limitations, we have posted our analysis on github (see data availability statement), allowing estimates to be calibrated with different risk factors, HRs and data from other countries, as more research on the virus emerges.

Twitter Paul Novosad @paulnovosad

Contributors All authors had full access to all the data in the study and shared the final responsibility for the decision to submit for publication. All authors saw and approved the final version of the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors were equally involved in every part of the study, including conception, analysis and writing. PN, RJ, AC and SA were involved in the paper conception, data analysis, interpretation and writing.

Funding Asher and Novosad received funding from Emergent Ventures, with application number #466.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was exempted from human subjects review by the Institutional Review Board at Dartmouth College.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Paul Novosad http://orcid.org/0000-0001-8331-6630
Radhika Jain http://orcid.org/0000-0002-4976-2167
Alison Campion http://orcid.org/0000-0001-9820-6450
Sam Asher http://orcid.org/0000-0002-2584-8011

REFERENCES

1 World Health Organization. *Coronavirus disease (COVID-19) Situation Report - 142*, 2020.
2 Onder G, Rezza G, Brusaferro S. Case-Fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* 2020;323:1775–6.
3 Dowd JB, Andriano L, Brazel DM, *et al*. Demographic science AIDS in understanding the spread and fatality rates of COVID-19. *Proc Natl Acad Sci U S A* 2020;117:9696–8.
4 Walker PGT, Whittaker C, Watson OJ, *et al*. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* 2020;369:eabc0035.
5 Oliveira MHSde, Wong J, Lippi G, *et al*. Analysis of clinical and demographic heterogeneity of patients dying from COVID-19 in Brazil versus China and Italy. *Braz J Infect Dis* 2020;24:273–5.
6 Updates on COVID-19: may 21 2020. Ministry of health and family welfare, government of India, 2020. Available: https://pib.gov.in/PressReleseDetailm.aspx?PRID=1625744 [Accessed 22 Jun 2020].
7 McCoy T, Traiano H. In the developing world, the coronavirus is killing far more young people, 2020. Available: https://www.washingtonpost.com/world/the_americas/coronavirus-brazil-killing-young-developing-world/2020/05/22/f76d83e8-99e9-11ea-ad79-eef7cd734641_story.html [Accessed 22 Jun 2020].
8 Laxminarayan R, Wahl B, Dudala SR, *et al*. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* 2020;370:691–7.
9 Minu P, Ray D, Subramanian S. *Decoding India's Low Covid-19 Case Fatality rate*. NBER Work Pap, 2020.
10 Ioannidis JPA, Axfors C, Contopoulos-Ioannidis DG. Population-Level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environ Res* 2020;188:109890.
11 Gilbert M, Pullano G, Pinotti F, *et al*. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *Lancet* 2020;395:871–7.
12 Pearson CAB, van Zandvoort K, Jarvis CI, *et al*. *Modelling projections for COVID19 epidemic in India. LSHTM CMMID COVID-19 Working group*, 2020.
13 Richardson S, Hirsch JS, Narasimhan M, *et al*. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the new York City area. *JAMA* 2020;323:2052–9.
14 Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020;323:1239–42.
15 Zhou F, Yu T, Du R, *et al*. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
16 Davies NG, Klepac P, Liu Y, *et al*. Age-Dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med* 2020;26:1205–11. doi:10.1038/s41591-020-0962-9
17 van Zandvoort K, Jarvis CI, Pearson CAB, *et al*. Response strategies for COVID-19 epidemics in African settings: a mathematical modelling study. *BMC Med* 2020;18:324. doi:10.1186/s12916-020-01789-2
18 Banerjee A, Pasea L, Harris S, *et al*. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. *Lancet* 2020;395:1715–25.
19 Cabore JW, Karamagi HC, Kipruto H, *et al*. The potential effects of widespread community transmission of SARS-CoV-2 infection in the world Health organization African region: a predictive model. *BMJ Glob Health* 2020;5:e002647.
20 Ghisolfi S, Almås I, Sandefur JC, *et al*. Predicted COVID-19 fatality rates based on age, sex, comorbidities and health system capacity. *BMJ Glob Health* 2020;5:e003094.
21 Clark A, Jit M, Warren-Gash C, *et al*. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *Lancet Glob Health* 2020;8:e1003–17.
22 Williamson EJ, Walker AJ, Bhaskaran K, *et al*. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020;584:430–6.
23 Petrilli CM, Jones SA, Yang J, *et al*. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 2020;369:m1966.
24 Lifestyles Team, NHS Digital. Health survey for England 2018, 2019. Available: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2018
25 World Health Organization. *World Health organization and international diabetes fund. definition and diagnosis of diabetes mellitus and intermediate hyperglycemia*, 2006. https://www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes_new.pdf
26 Global Burden of Disease Collaborative Network. *Global burden of disease study 2017 (GBD 2017) results*. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018. http://ghdx.healthdata.org/gbd-results-tool
27 McLean S, Hoogendoorn M, Hoogenveen RT, *et al*. Projecting the COPD population and costs in England and Scotland: 2011 to 2030. *Sci Rep* 2016;6:31893.
28 Leung JM, Sin DD. Smoking, ACE-2 and COVID-19: ongoing controversies. *Eur Respir J* 2020;56. doi:10.1183/13993003.01759-2020. [Epub ahead of print: 16 Jul 2020].
29 Cummings MJ, Baldwin MR, Abrams D, *et al*. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *medRxiv* 2020;395:1763–70. doi:10.1016/S0140-6736(20)31189-2
30 Das J, Sánchez-Páramo C. Smart containment: how low-income countries can tailor their COVID-19 response, 2020. Available: https://blogs.worldbank.org/voices/smart-containment-how-low-income-countries-can-tailor-their-covid-19-response [Accessed 20 Jun 2020].
31 Covid19 tracker- fatalities: top 10 comorbidities by age group, 2020. Available: https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Fatalities?%3

The COVID-19 mortality effects of underlying health conditions in India: a modelling study Appendix

Paul Novosad*

Radhika Jain†

Alison Campion‡

Sam Asher§

July 2020

1 Appendix

1.1 Construction of Indian survey dataset

The 4th District Level Health Survey (DLHS-4) and Indian Annual Health Survey (AHS) were conducted between 2012 and 2014 with no overlap in geographic coverage and jointly cover all states and union territories in India except Jammu and Kashmir, Dadra and Nagar Haveli, and Lakshadweep (additionally, data for Gujarat were collected but not made publicly available). The DLHS-4 is a single cross-sectional survey, while the AHS is a three year panel survey. Both surveys are representative at the district level and use two-stage stratified cluster sampling. In rural areas the primary sampling unit (PSU) was a village and in urban areas the PSU was a census enumeration block in the AHS or an urban frame survey block in the DLHS-4. PSUs were selected randomly with probability proportional to population size using the 2001 Indian Population Census in the AHS and rural DLHS-4, and with equal probability in the urban DLHS-4. Households were the secondary sampling unit (SSU) and were selected through systematic random sampling [1].

The DLHS-4 and AHS surveys administered a household-level questionnaire that collected information on age, sex, and self-reported symptoms, diagnosis, and treatment of illness for each household member. Additionally, both surveys administered a Clinical, Anthropometric and Biochemical (CAB) module to collect data on height, weight, hemoglobin, blood pressure, and blood glucose for adults. The CAB module was completed for all individuals 18 years or older in all sampled households in the DLHS-4. In the AHS, the CAB module was conducted in the second round of the survey in 2014 for all individuals 18 years or older in a randomly selected subsample of twelve PSUs per district, on average [2]. The publicly available DLHS-4 data provides a merged dataset in which each individual in the CAB module is matched to their household survey responses. However, there is no merged AHS data available, and there is no unique identifier to merge individuals between the household and CAB survey modules. We merged individuals with a completed CAB module in the AHS survey to their records in the household module using state, district, stratum (urban and non-urban), household unit, household number, and individual serial number identifiers. Individuals missing one or more of these key identifying fields could not be uniquely identified or merged and were dropped from the analysis. Of 1,209,926 individuals covered in the CAB module of the AHS, 819,351 (67.7%) had all required identifying fields and were matched with their records in the household survey.

We combined the DLHS-4 and AHS datasets to create a nationally representative dataset for analysis. Individuals with missing age, sex, height, weight, glucose, or blood pressure measurement were dropped from the sample. Individuals with reported age greater than 99 were assumed to be outliers and also dropped from the dataset. Additionally, as the thresholds for defining risk factors like obesity during pregnancy are not well established, pregnant women were also dropped from the sample. The final analysis sample contained 1,375,548 individuals, of which 577,994 (42.0%) came from the AHS and the remaining 797,554 (58.0%) came from the DLHS.

1.2 Estimation of glucose, obesity, and blood pressure in India

The DLHS-4 and AHS surveys both used the same data collection methods for biomarkers. Details of biomarker measurement are described in the survey manuals and summarized briefly here [2]. Systolic and diastolic BP were measured twice on the upper left arm, while sitting, with an interval of at least 3 minutes for each individual. The mean of the two measures was used to generate a single continuous measure of BP that was then classified as hypertension if systolic BP was 140 mmHg or higher or diastolic BP was 90 mmHg or higher, or if the individual reported a diagnosis of hypertension. Height and weight were directly measured and BMI was calculated as weight in kilograms divided by the square of height in meters. Blood glucose was measured from a single capillary blood sample (finger prick) and automatically converted into plasma equivalents by the glucometer. Single

*Dartmouth College
†Stanford University
‡Development Data Lab
§Johns Hopkins University School of Advanced International Studies

Novosad P, *et al. BMJ Open* 2020; 10:e043165. doi: 10.1136/bmjopen-2020-043165

capillary glucose measures are not ideal for clinical diagnosis of diabetes but have been recommended by the WHO for population surveillance in lower income countries [3]. Standard international thresholds were used to define diabetes as a plasma glucose reading ≥126mg/dL [7.0mmol/L] if fasting or ≥200 mg/dL [11.1 mmol/L] if reported not fasting. Individuals were asked to fast overnight before their glucose measurement. Self-reported fasting status was recorded in the DLHS-4 but not the AHS. We follow other studies that have used these data and use self-reported fasting status for all DLHS-4 participants and assume all AHS participants had fasted for the primary analysis [4, 5]. Assuming these individuals were not fasting changes estimated total diabetes prevalence in India from 9.8% to 8.5%.

All prevalence estimates from the DLHS-4 and AHS were weighted with a sample weight. Sample weights determined by the survey design of the DLHS-4 were provided in the publicly available data. These weights were multiplied by a district population weight, defined as the percentage of the national population in each district, to obtain the final sample weight. Due to a different survey design, the AHS does not have a sample weight in the data and so the sample weight was defined only by the district population weight.

1.3 Age-specific prevalence of risk factors in India and England

Table A1 shows age-specific prevalence of each risk factor considered in the study. Definitions of risk factors and data sources are given in the methods section and in Section 1.2.

Table A1
Health condition prevalences

	Prevalence (%)											
	India						**England**					
Age Group	18–39	40–49	50–59	60–69	70–79	80–99	18–39	40–49	50–59	60–69	70–79	80–99
Diabetes (Controlled)	0.3	1.7	3.1	3.9	4.3	3.3	0.7	5.0	8.8	11.3	13.9	13.9
Diabetes (Uncontrolled)	4.8	10.1	14.0	15.7	16.3	16.3	0.3	1.8	3.2	3.6	4.0	4.0
Hypertension	16.8	31.3	40.3	47.5	52.3	52.5	6.8	18.3	32.0	49.2	61.3	66.0
Obese (class I & II)	3.0	5.5	5.6	4.8	3.6	2.3	17.7	27.6	30.9	30.0	27.9	26.4
Obese (class III)	0.4	0.5	0.5	0.5	0.3	0.3	2.6	3.8	4.1	3.6	2.7	1.7
Chronic Heart Disease	1.4	4.4	8.1	15.0	24.5	31.4	1.3	5.0	10.5	21.5	34.7	42.6
Chronic Respiratory Disease	1.2	4.6	10.6	19.0	25.7	27.5	0.0	0.5	1.8	4.8	8.2	9.3
Asthma	1.3	2.7	4.2	6.2	8.1	8.6	9.3	8.5	8.4	8.5	8.4	7.8
Kidney Disease	6.6	13.1	17.3	24.4	37.0	51.7	3.0	5.5	7.9	14.2	27.2	45.8
Chronic Liver Disease	5.5	5.9	5.9	5.7	5.5	5.1	2.6	3.6	3.7	3.7	3.5	3.3
Haematological Cancer	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.2	0.4	0.7	1.0	1.0
Non-haematological Cancer	0.1	0.4	0.8	1.0	1.1	1.2	1.1	2.7	4.6	8.0	11.3	12.6
Stroke, Dementia	0.2	1.0	2.2	4.2	8.0	14.7	0.2	0.8	1.9	4.2	10.4	23.0
Other Neurological Condition	0.1	0.1	0.1	0.0	0.0	0.0	0.2	0.2	0.2	0.1	0.0	0.0
Psoriasis, Rheumatoid	0.8	1.4	1.8	2.3	2.5	2.3	2.0	3.4	4.5	5.2	4.8	3.7
Other Immunosuppressive Conditions	0.2	0.3	0.2	0.1	0.0	0.0	0.1	0.3	0.2	0.2	0.1	0.0

1.4 Demographics and prevalence of health conditions in England and the OpenSAFELY study population

We obtained estimated COVID-19 mortality hazard ratios for risk factors from the OpenSAFELY study [6]. The OpenSAFELY study sample includes adults 18 years or older enrolled with The Phoenix Partnership general practice system in England and covers 40% of the English population.

In order to calculate population risk for England for this study, we obtained national age, sex, and risk factor prevalence for the entire English population from a combination of census data, population health surveys and the GBD, as described in the methods. In Table A2 we present characteristics of the OpenSAFELY study sample against those of the entire English population as represented in our population statistics. Age, sex, and prevalence of most risk factors are very similar in the two. Hypertension prevalence in the OpenSAFELY sample is higher than in the English population (34.2% vs 28.1%) and asthma and class I obesity are lower (1.7% vs 9.2% and 19.1% vs 24.8%).

Table A2

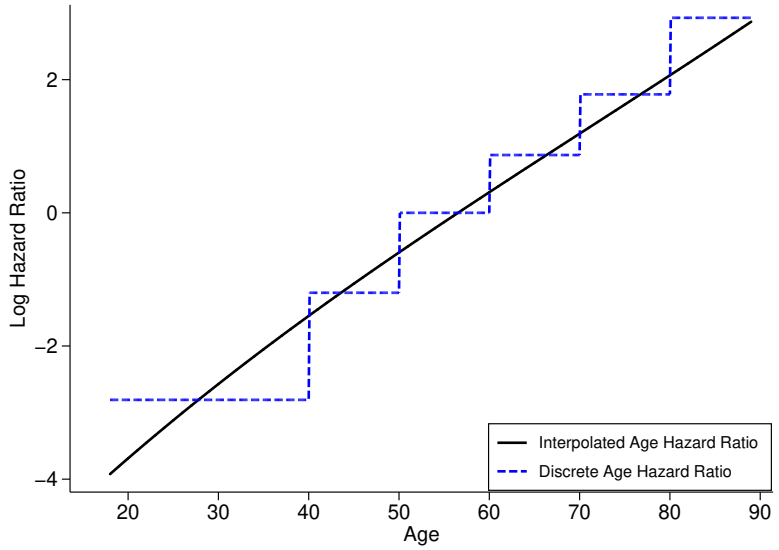
Demographics and prevalence of conditions in England and in OpenSAFELY

	England Prevalence (%)	
	OpenSafely Sample	This Study
Age 18-39	34.4	36.6
Age 40-49	16.5	16.3
Age 50-59	17.6	17.0
Age 60-69	13.8	13.3
Age 70-79	11.2	10.4
Age 80-99	6.5	6.3
Male	49.9	48.9
Diabetes (Controlled)	6.0	6.4
Diabetes (Uncontrolled)	2.8	2.1
Hypertension	34.2	28.0
Obese (class I & II)	19.1	24.8
Obese (class III)	2.7	3.1
Chronic Heart Disease	6.7	5.9
Chronic Respiratory Disease	4.1	2.5
Asthma	1.7	9.2
Kidney Disease	6.3	5.6
Chronic Liver Disease	0.7	2.6
Haematological Cancer	0.1	0.2
Non-haematological Cancer	0.5	2.6
Stroke, Dementia	2.1	1.5
Other Neurological Condition	1.0	0.1
Psoriasis, Rheumatoid	5.1	2.4
Other Immunosuppressive Conditions	1.6	0.1

1.5 Interpolation of age bin relative risks

Hazard ratios for age in OpenSAFELY are reported in the discrete bins 18–39, 40–49, 50–59, 60–69, 70–79, 80+. To obtain hazard ratios at continuous ages, we first converted hazard ratios to natural logs and then fitted a cubic polynomial to the midpoints of each bin. The hazard ratio is almost linear in age and the polynomial provides a very good fit (Figure A1). These continuous log hazard ratios were converted into relative risks assuming a mortality rate of 1% (as described in the methods section), and then collapsed into integer age bins for the analysis.

Figure A1
Age Interpolation: Fully-Adjusted Model



1.6 Sensitivity of results to sampling error

The two sources of sampling error in the data underlying the analysis are the hazard ratios in OpenSAFELY (which are reported with 95% confidence intervals) and the health condition prevalence estimates. We examined sensitivity to sampling error by resampling from simulated datasets with distributions indicated by the standard errors in the underlying data.

To examine sensitivity to sampling error in hazard ratios, we ran the analysis on 1000 samples where each hazard ratio was replaced with a number drawn from a distribution with mean and standard deviation equal to the reported hazard ratio and calculated standard error from OpenSAFELY. Draws were conducted in logs and then converted back to levels and relative risks for the analysis. We examined the distribution of estimates for: (A) the combined population relative risk from health conditions in India relative to England; (B) the modeled share of deaths under age 60 in England; and (C) the modeled share of deaths under age 60 in India (Figure A2). The red line in the figure indicates the statistic reported in the results.

In each case, the entire distribution of results is highly consistent with what is reported in the results. Population relative risk due to underlying health conditions is consistently lower in India than in England, with a 95% confidence interval of [0.86,0.97]. Nearly all draws for the share of deaths under 60 in England and in India are within half a percent of the primary estimate.

We performed a similar exercise to examine sensitivity to sampling error in prevalence estimates (Figure A3). For biomarker data from India, standard errors were calculated directly from the data. For prevalences obtained from GBD and for the England biomarkers, we used the standard errors or 95% confidence intervals reported with the underlying data.

The sensitivity of the results to sampling error in prevalences is even smaller than the sensitivity with respect to hazard ratios, likely because prevalence estimates are based on very large samples, while hazard ratios are calculated from only a small number of deaths. The 95% confidence interval for population relative risk in this simulation was [0.88,0.93], and the percentage of deaths under 60 in both countries also showed little variation.

Figure A2
Sensitivity Test 1: Hazard Ratio Uncertainty

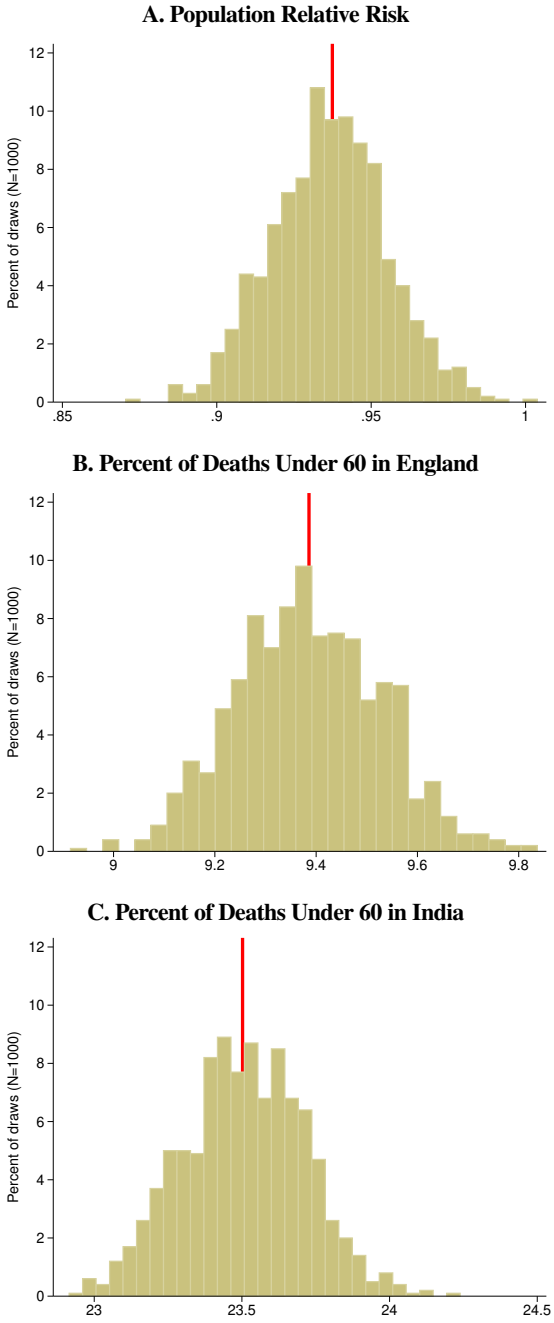
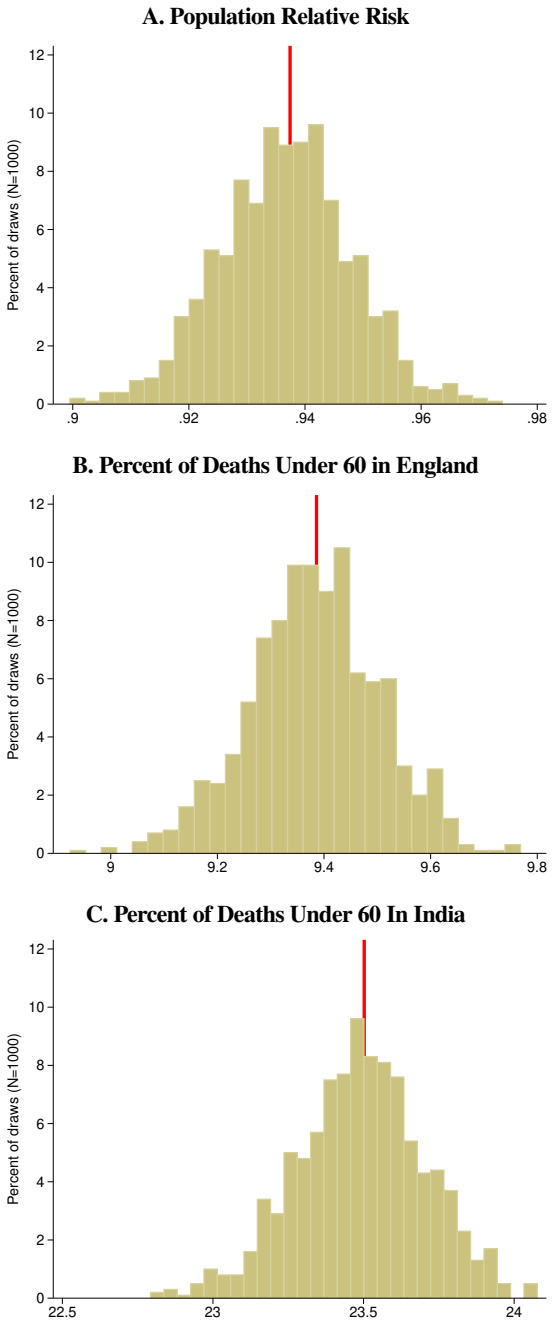


Figure A3
Sensitivity Test 2: Prevalence Uncertainty



1.7 Sensitivity to correlated health conditions

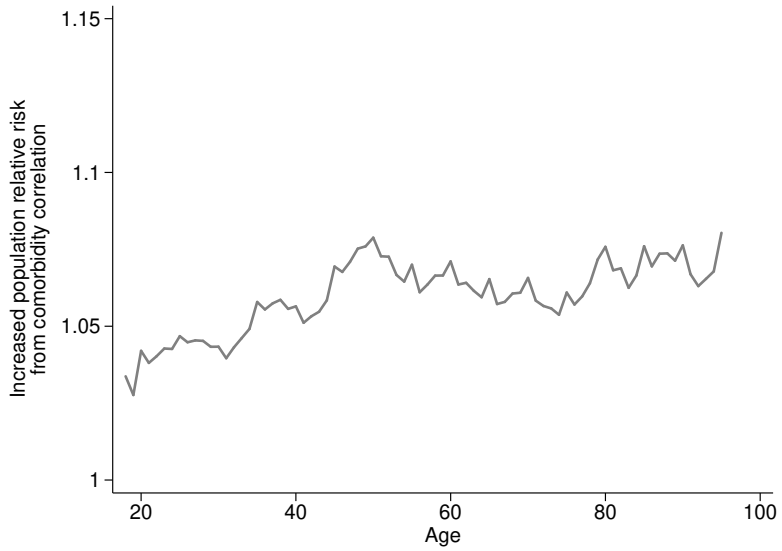
Our primary analysis implicitly assumes that COVID-19 mortality risk factors are uncorrelated with each other. This assumption is unavoidable given that many of the risk factors in India are drawn from the Global Burden of Disease studies which report age-specific prevalence but not the correlation across conditions.

If risk factors are correlated, then actual mortality risk will be higher than what is estimated because the relative risks of joint comorbidities will be compounded. Even if mortality risk is biased downward, our results may not be substantially affected because they rely on comparisons between England and India rather than focusing on the relative risk directly. If England and India have similar covariances of health conditions, then the biases will balance out exactly. Note that our analysis accounts for the most significant correlation of risk factors, which is the positive correlation between health conditions and age. This is accounted for because all health conditions are aggregated at each age.

To study the impact of correlated health conditions on population relative risk, we focused on the subset of health conditions available in the Indian microdata: obesity, diabetes, and hypertension. We calculate individual relative risk from the combined set of these risk factors by compounding the relative risks for these conditions at the individual level. We then aggregated the data across individuals and compared the combined relative risk to the same measure generated from the aggregate data, i.e. without accounting for correlation between comorbidities (Figure A4). The covariance of conditions increases the mortality rate by 5–10% across the age distribution. The effect does not vary much by age and is small relative to the aggregate risk factor shown in Figure 2.

The close match of our model to the age-specific death rates in OpenSAFELY is a second indication that condition covariance is unlikely to substantially bias the findings. As noted in the body of the paper, our model predicts 8.8% of deaths below the age of 60 in England; OpenSAFELY reports a figure of 8.6%. Because our hazard ratios come from an individual model that takes correlation between conditions into account, the close match of the age distribution of deaths suggests that our inability to observe condition covariance has not biased the results substantially.

Figure A4
Sensitivity Test 3: Covariance of Health Conditions



References

[1] Office of the Registrar General & Census Commissioner GoI India Ministry of Home Affairs. Annual Health Survey Report: A Report on Core and Vital Health Indicators Part I; 2014.

[2] Office of the Registrar General & Census Commissioner GoI India Ministry of Home Affairs. Annual Health Survey Report: A Report on Clinical, Anthropometric and Bio-Chemical Survey Part II. Office of the Registrar General & Census Commissioner, India Ministry of Home Affairs, Government of India; 2014.

[3] Organization WH, Fund ID. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. World Health Organizatoin; 2006.

[4] Geldsetzer P, Manne-Goehler J, Theilmann M, Davies JI, Awasthi A, Vollmer S, et al. Diabetes and Hypertension in India: A Nationally Representative Study of 1.3 Million Adults. JAMA Internal Medicine. 2018;178(3):363–372.

[5] Bischops AC, Manne-Goehler J, Jaacks LM, Awasthi A, Theilmann M, Davies JI, et al. The prevalence of concurrently raised blood glucose and blood pressure in India: a cross-sectional study of 2,035,662 adults. Journal of Hypertension. 2019;37(9):1822–1831.

[6] Williamson E, Walker AJ, Bhaskaran KJ, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. medRxiv. 2020:2020.05.06.20092999. Available from: `http://medrxiv.org/content/early/2020/05/07/2020.05.06.20092999.abstract`.