medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

All-cause excess mortality in the State of Gujarat, India, during the COVID-19 pandemic (March 2020-April 2021)

Rolando J. Acosta[1], Biraj Patnaik[2], Caroline Buckee[3,4], Satchit Balsari[*5,6], Ayesha Mahmud[*7]

[1]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02115
[2]National Foundation for India, New Delhi 110 003, India
[3]Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, MA 02115
[4]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115
[5]Emergency Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA
[6]François-Xavier Bagnoud Center for Health and Human Rights, Harvard T. H. Chan School of Public Health, Boston, MA 02115
[7]Department of Demography, University of California, Berkeley, CA 94720
*balsari@hsph.harvard.edu
*mahmuda@berkeley.edu

Abstract

Official COVID-19 mortality statistics are strongly influenced by the local diagnostic capacity, strength of the healthcare system, and the recording and reporting capacities on causes of death. This can result in significant undercounting of COVID-19 attributable deaths, making it challenging to understand the total mortality burden of the pandemic. Excess mortality, which is defined as the increase in observed death counts compared to a baseline expectation, provides an alternate measure of the mortality shock of the COVID-19 pandemic. Here, we use data from civil death registers for 54 municipalities across the state of Gujarat, India, to estimate the impact of the COVID-19 pandemic on all-cause mortality. Using a model fit to monthly data from January 2019 to February 2020, we estimate excess mortality over the course of the pandemic from March 2020 to April 2021. We estimated 16,000 [95% CI: 14,000, 18,000] excess deaths across these municipalities since March 2020. The sharpest increase in deaths was observed in April 2021, with an estimated 480% [95% CI: 390%, 580%] increase in mortality from expected counts for the same period. Females and the 40 to 60 age groups experienced a greater increase from baseline mortality compared to other demographic groups. Our excess mortality estimate for these 54 municipalities, representing approximately 5% of the state population, exceeds the official COVID-19 death count for the entire state of Gujarat.

1 Introduction

Official COVID-19 mortality counts, around the world, largely rely on the attribution of COVID-19 as a cause of death on death certificates[1, 2, 3]. Data from death certificates is then aggregated to centralized databases, often with some reporting delay[4, 5, 6]. These data can be assumed to be sensitive and reliable indicators of the true national toll of COVID-19 related deaths when most patients with COVID-19 have access to healthcare, health care providers have the knowledge and tools necessary to diagnose COVID-19, and those recording deaths on death certificates have the requisite training to note COVID-19 as an underlying cause of death on the certificate[2, 7]. One or more of these conditions is often not met, resulting in underestimates of COVID-19 related deaths, as is the case in India[8].

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

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

In the absence of reliable death registration data in the aftermath of disasters and public health emergencies, scientists have relied on alternative methods to estimate deaths, including through house-hold based surveys, crematoria and funeral home body counts, or verbal biopsies[9, 10, 11, 12]. In many countries, the estimation of all-cause mortality has provided an alternative proxy for the underestimation of COVID-19 attributable deaths in official statistics[13, 14, 15]. During disasters and public health crises, all-cause mortality estimates can provide an overall measure of the mortality shock, including deaths resulting directly from the disaster, as well as those resulting from indirect impacts of the disaster, like disrupted access to health care[16]. During the COVID-19 pandemic, all-cause mortality data include both directly attributable deaths (those that died from SARS-Cov-2 and its complications) and indirectly attributable deaths (those that died from the indirect impacts of the pandemic, including delayed or deferred care for other conditions)[17]. While all-cause mortality estimates have been compared to expected all-cause mortality based on historical data to understand the true toll of the pandemic[14, 18, 19], few studies have looked at this excess mortality in a low-income setting. News reports and lived experience suggests that the true pandemic-related death toll in India is larger than the official estimate. Here, we use deaths data from multiple municipalities in the State of Gujarat in India to examine the impact of the COVID-19 pandemic of all-cause mortality; we explore differences in death counts by age and sex over the course of the pandemic, and estimate excess mortality resulting from the pandemic.

2 Methods

We used de-identified data from civil death registers, the official records at the municipal jurisdictional level, where deaths are first recorded in the death registers maintained at the Gram Panchayats (village committees) in rural India, and in municipalities and municipal corporations in urban India. A few-weeks lag is normal at this stage. Thereafter, the data are then aggregated at the district level and submitted to the national Civil Registration System (CRS)[20, 21]. It may take up to nine months from the end of the financial year for the CRS to be fully updated and validated by the government of India. Therefore, data collated directly from the death registers maintained locally, are the most comprehensive source of deaths from 2020 and 2021 currently available - until the CRS is fully updated. We conducted a secondary analysis of data directly derived from death registers from 54 (of 162) municipalities across 24 (of 33) districts in the state of Gujarat, representing a population of at least around 3.2 million (according to the 2011 census) , or approximately 5% of the total state population of 69 million[22]. These data were procured by The Reporters' Collective, a group of investigative journalists in India, and made publicly available[23, 24]. These data encompass all recorded deaths from January 2019 to April 2021, and include date of death, date of registration, gender, age, and place of death information.

In Gujarat, according to the National Family Health Survey (NFHS) 2019-20, 93% of all deaths "of usual residents of households" in Gujarat are recorded in the civil death registers. Death registration completeness reaches nearly 96% in urban areas, 92% in rural areas, 94% for males, and 91% for females, overall[25]. Using these records, we computed monthly mortality counts stratified by 10-year age groups, gender, and place of death from January 2019 to April 2021.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2.1 Statistical Analysis

Based on the reported COVID-19 case and mortality data in India from early 2020, we considered data from January 2019, to February 2020 to represent baseline mortality. We compared the baseline mortality data to observed counts from March 2020 onwards to estimate excess mortality. First, we aggregated the data across all demographic indicators to obtain monthly death counts for all selected municipalities in the state. Let Y_t represent the number of deaths at month t and assume that $Y_t \sim \text{Poisson}(\mu_t)$, where:

$$\log(\mu_t) = \alpha + \beta t + s(m_t) \text{ for } t = 1, \ldots, T \tag{1}$$

In model (1), μ_t represents the average number of deaths at month t, α is an intercept term, β is a linear effect of time that accounts for a non-seasonal trend, s is a harmonic component that accounts for seasonal mortality, $m_t \in \{1, \ldots, 12\}$ corresponds to the month of the year associated with t, and T is the total number of observations. Due to the unavailability of reliable population estimates, we assumed that the population size in the state was constant from 2019 to 2021, and we did not include a population offset in our mean model. We used data from January 2019 to February 2020 to fit model (1) via maximum likelihood assuming an overdispersed Poisson distribution. Then, we estimated expected mortality from March 2020 to April 2021 using the estimated model parameters. To account for variation in the observed counts during the period of interest, we let t' be a month after February 2020 with corresponding count $Y_{t'} \sim \text{Poisson}(\mu_{t'})$, where:

$$\log(\lambda_{t'}) = \gamma + f(t') + \log(\hat{\mu}_{t'}) \tag{2}$$

, with $\lambda_{t'}$ the average number of deaths at t', γ an intercept, f a smooth function of time that represents deviations from expected mortality based on historical data, and $\log(\hat{\mu}_{t'})$ an offset representing expected mortality. We modeled f with a natural cubic spline with 11 internal knots and, as before, fit model (2) via maximum likelihood assuming an overdispersed Poisson distribution. We estimated excess deaths at t' with:

$$\hat{\delta}_{t'} = \hat{\lambda}_{t'} - \hat{\mu}_{t'} \tag{3}$$

, with variance

$$\text{Var}\left(\hat{\delta}_{t'}\right) = \hat{\phi}_\lambda \hat{\lambda}_{t'} - \hat{\phi}_\mu \hat{\mu}_{t'} \tag{4}$$

, where $\hat{\phi}_\mu$ and $\hat{\phi}_\lambda$ are the estimated dispersion parameters from models (1) and (2), respectively. We estimated cumulative excess death and associated confidence intervals by summing the excess death estimates and corresponding variance estimates, respectively. We use the `excessmort` R package for this analysis[26]. In the Results section, we round our excess death estimates proportional to the standard error. For example, if the standard error is in the hundreds, then we round our excess death estimate to the nearest tens.

3 Results

3.1 Observed deaths overall and by age and sex

44,568 total deaths were recorded across the 54 municipalities over the course of the pandemic, from March 2020 onwards. While deaths were higher in both 2020 (31,477) and 2021 (17,882 up to April) compared to 2019 (25,590), the sharpest increase in deaths was observed during the second wave of the pandemic in 2021 (Figure 1). Between January and April of

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

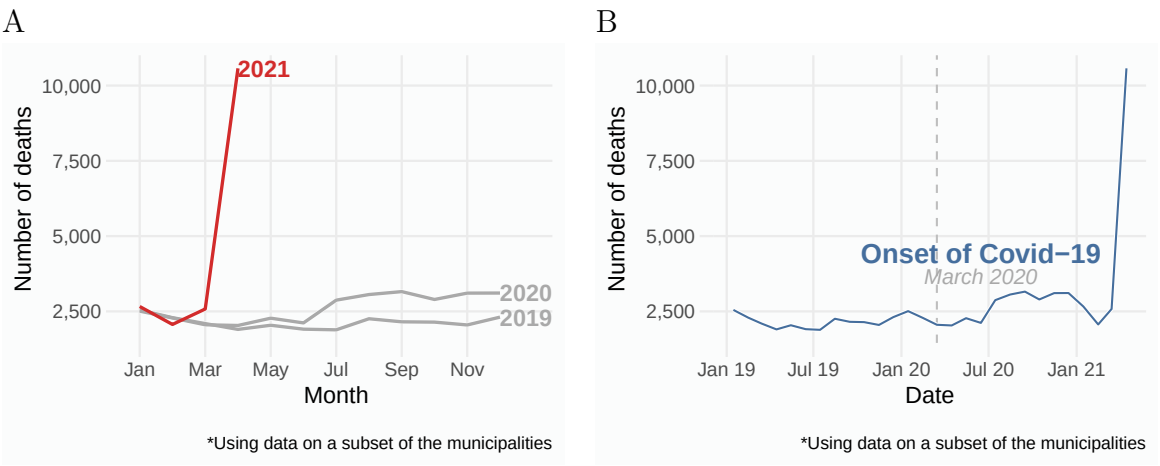

Figure 1: Monthly deaths counts in Gujarat, India, from January 2019 to April 2021. A) The gray curves represent the observed monthly counts in 2019 and 2020, respectively. The red curve represents the observed monthly counts from January to April 2021. B) The same data as in A) but now shown as a contiguous time series. The vertical dashed-gray line corresponds to the onset of Covid-19 in Gujarat as we defined it.

2021, 17,882 deaths were observed, reflecting a 102% increase over the average of the previous two years for the same months. The observed increase in all-cause mortality between January and April 2021 differed by age and sex. The largest percentage change compared to the same months in the previous two years was in the 50 to 60 years age group (164%) followed by the 40 to 50 years age group (152%). Conversely, the smallest percentage change was in the 10 and under age group (-22%). Females experienced a slightly larger increase in mortality in 2021 (107%) than males (103%), although males had higher mortality counts throughout the entire study period. The gender discrepancy was particularly acute in the 40 to 50 and 50 to 60 years age group, where we found a difference of 91 percentage points in increased mortality between females and males in the 40 to 50 years age group and a difference of 79 percentage points in the 50 to 60 years age group (Figure 2).

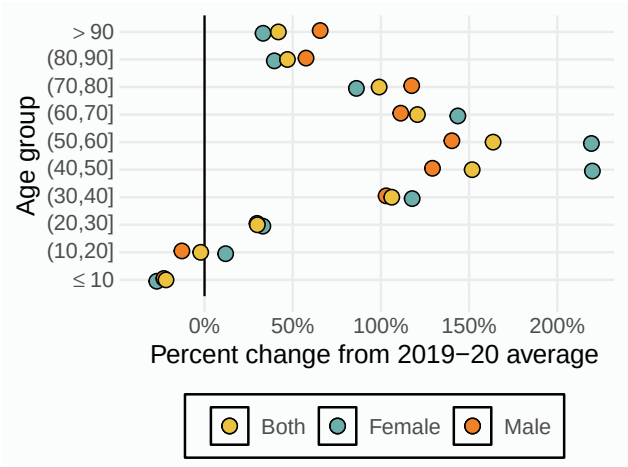

Figure 2: Percent change in total deaths from January to April 2021 relative to the average of 2019-20 for the same period, stratified by age groups and gender.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

3.2 Excess deaths during the pandemic

We estimated 16,000 [95% CI: 14,000, 18,000] excess deaths across the 54 municipalities in Gujarat since March 2020. However, most of these deaths occurred during the second wave, between January and April 2021, with 9,500 [95% CI: 7,700, 11,300] estimated excess deaths. The most striking deviation from baseline is for the month of April 2021, where we estimate 480% [95% CI: 390%, 580%] more deaths than expected (Figure 3).

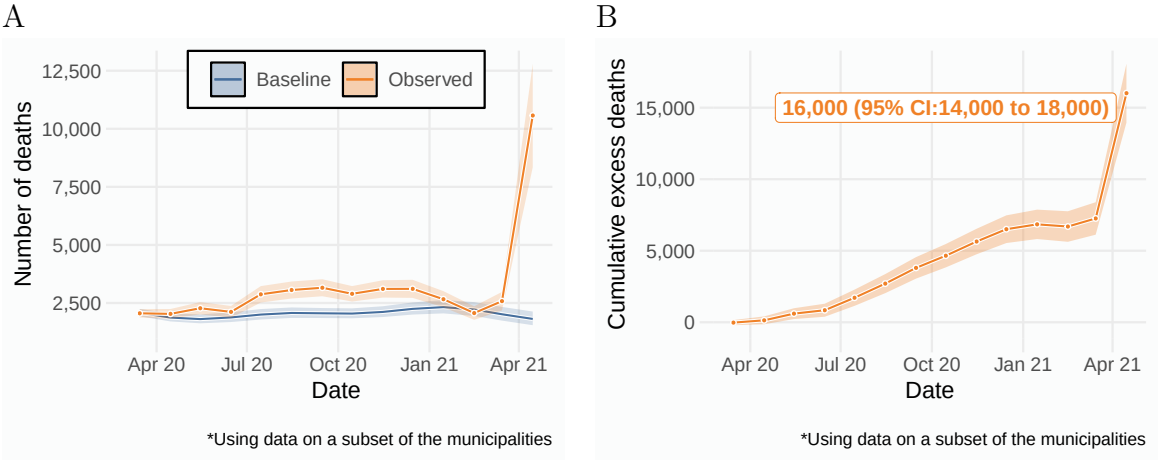

Figure 3: Excess deaths in Gujarat, India, from March 2020, to April 2021. A) Expected and observed monthly counts from March 2020, to April 2021. The blue curve and shaded region represent the average number of deaths and pointwise 95% confidence interval based on historical data, respectively. The orange curve represents the recorded monthly counts for the same period and the orange shaded region represents a 95% confidence interval obtained from model (2). B) Cumulative excess deaths from March 2020, to April 2021. The orange curve and shaded region represent our cumulative excess deaths estimate and pointwise 95% confidence interval, respectively.

4 Discussion

We describe mortality trends across 54 municipalities in Gujarat, India over the course of the COVID-19 pandemic, until April 2021. The official death count for the entire state of Gujarat from March 2020 to August 16 2021 is 10,075[27]. Our results suggest that in these 54 municipalities alone there were 16,000 [95% CI: 14,000, 18,000] excess deaths from March 2020 to April 2021, with females and the 40 to 60 age groups experiencing a greater increase from baseline mortality compared to other demographic groups. The vast majority of these excess deaths likely represent direct deaths from COVID-19, in the absence of any other known catastrophe. A small percentage of these would include deaths from the indirect impact of the pandemic, and from causes unrelated to the pandemic.

Our study has several limitations. First, the data only represent around 5% of the population of Gujarat covering 54 of 162 municipalities[22]. These are urban municipalities. With the exception of Gandhinagar, they do not include data from the municipal corporations of other large urban centers, as these data were unavailable. They also do not include data from the rural gram panchayats. Though the municipalities were spread across the state (see Supplement 1) they represent a convenience sample rather than a random sample, and we are unable to extrapolate our results to estimate deaths across the entire state. Given the high percentage of deaths recorded in the registers, per the NFHS, these data are, however,

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

highly representative of mortality in the municipalities examined. The strikingly high mortality is also consistent with media reports and lived experience and likely representative of the general trend across the state.

Second, since we had no data on the yearly population size for each municipality we were unable to calculate mortality rates and make comparisons across municipalities. For the same reason, we were unable to assess excess mortality by demographic indicators. The last published census data are from 2011. While data from electoral rolls are more recent, they do not map to the same geospatial unit of the municipality, and cannot be easily used. We therefore make the assumption that the population remained unchanged between January 2019 and April 2021. Our results will be biased if there was significant migration in or out of the state between our baseline period and the pandemic period. Had population sizes significantly increased (or decreased) over time, our excess mortality estimate would be an overestimate (or underestimate). Because mortality varies strongly with age, both for COVID-19 and all-cause mortality, an ideal comparison would also adjust for age-specific population changes.

Third, we only have baseline (pre-pandemic) data from January 2019 to February 2020. Since the baseline period for fitting the model is relatively short, we may not be sufficiently capturing year-to-year variations in mortality. However, the sharp increase in mortality observed in 2021 is unlikely to fall within the bounds of normal yearly variability in mortality. Finally, there may be lags in recording of deaths in the death registry, and not all deaths may yet be registered. According to media reports, mortality continued to be high, or rose, in May 2021, and is not yet included in the published data or in our estimates.

We estimated a 480% increase [95% CI: 390%, 580%] in deaths in April 2021, in the municipalities studied. This is the highest percentage increase in deaths recorded in a single month anywhere in the world. In April 2020, Ecuador recorded a 411% increase; in April 2021, Peru recorded a 345% increase[28]. This large discrepancy between official COVID-19 death counts and excess mortality underscores the need to rectify how official death counts are collated. Reliance on death certificates as the single source of truth is sub-optimal when access to health systems, testing availability, and death certification accuracy and completeness are all weak.

The high mortality counts across age groups warrants further investigation into the impact of underlying social determinants and the efficacy of clinical protocols and public health policies on mortality. The lack of relevant data precludes these necessary analyses. Globally, data on population estimates, testing, and clinical outcomes, where available, have facilitated contextually intelligent public health planning and response. State supported data transparency and availability can in fact help local scientists focus on knowledge generation, and provide citizens and the state the tools needed to strengthen health systems.

5 Acknowledgements

We thank The Reporters' Collective[23] for providing the un-identified data that was also made publicly available via a creative commons license at www.wallofgrief.org

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

References

[1] Daniel Howdon, Jason Oke, and Carl Heneghan. Death certificate data: Covid-19 as the underlying cause of death, September 2020. URL https://www.cebm.net/covid-19/death-certificate-data-covid-19-as-the-underlying-cause-of-death/.

[2] David Oliver. David oliver: Mistruths and misunderstandings about covid-19 death numbers. *bmj*, 372, 2021.

[3] Patrick Boyle. How are covid-19 deaths counted? it's complicated, February 2021. URL https://www.aamc.org/news-insights/how-are-covid-19-deaths-counted-it-s-complicated.

[4] Center for Disease Control and Prevention. Technical notes: Provisional death counts for coronavirus disease (covid-19). URL https://www.cdc.gov/nchs/nvss/vsrr/covid19/tech_notes.htm.

[5] The Economic Times. Covid death recount: Late reporting is not peculiar to india alone, June 2021. URL https://economictimes.indiatimes.com/news/india/covid-death-recount-late-reporting-is-not-peculiar-to-india-alone/articleshow/83451989.cms.

[6] Sara Simon, Ryan Briggs, and Nina Feldman. A long time to wait, September 2020. URL https://www.spotlightpa.org/news/2020/09/pa-coronavirus-deaths-reporting-edrs/.

[7] Guidance for certifying deaths due to coronavirus disease 2019 (covid–19), 2020.

[8] Shamika Ravi. Counting deaths in india is difficult, July 2021. URL https://www.hindustantimes.com/opinion/counting-deaths-in-india-is-difficult-101626273326958.html.

[9] Alfred Sommer and WileyH Mosley. East bengal cyclone of november, 1970: epidemiological approach to disaster assessment. *The Lancet*, 299(7759):1030–1036, 1972.

[10] Nishant Kishore, Domingo Marqués, Ayesha Mahmud, Mathew V Kiang, Irmary Rodriguez, Arlan Fuller, Peggy Ebner, Cecilia Sorensen, Fabio Racy, Jay Lemery, et al. Mortality in puerto rico after hurricane maria. *New England journal of medicine*, 379 (2):162–170, 2018.

[11] Peter Heudtlass, Niko Speybroeck, and Debarati Guha-Sapir. Excess mortality in refugees, internally displaced persons and resident populations in complex humanitarian emergencies (1998–2012)–insights from operational data. *Conflict and health*, 10(1): 1–11, 2016.

[12] Carlos Santos-Burgoa, John Sandberg, Erick Suárez, Ann Goldman-Hawes, Scott Zeger, Alejandra Garcia-Meza, Cynthia M Pérez, Noel Estrada-Merly, Uriyoan Colón-Ramos, Cruz María Nazario, et al. Differential and persistent risk of excess mortality from hurricane maria in puerto rico: a time-series analysis. *The Lancet Planetary Health*, 2 (11):e478–e488, 2018.

[13] The true death toll of covid-19: Estimating global excess mortality, September 2020. URL https://www.who.int/data/stories/the-true-death-toll-of-covid-19-estimating-global-excess-mortality.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

[14] Ariel Karlinsky and Dmitry Kobak. Tracking excess mortality across countries during the covid-19 pandemic with the world mortality dataset. *Elife*, 10:e69336, 2021.

[15] Thomas Beaney, Jonathan M Clarke, Vageesh Jain, Amelia Kataria Golestaneh, Gemma Lyons, David Salman, and Azeem Majeed. Excess mortality: the gold standard in measuring the impact of covid-19 worldwide? *Journal of the Royal Society of Medicine*, 113(9):329–334, 2020.

[16] Debarati Guha-Sapir and Francesco Checchi. Science and politics of disaster death tolls, 2018.

[17] Mathew V Kiang, Rafael A Irizarry, Caroline O Buckee, and Satchit Balsari. Every body counts: measuring mortality from the covid-19 pandemic. *Annals of Internal Medicine*, 173(12):1004–1007, 2020.

[18] Steven H Woolf, Derek A Chapman, Roy T Sabo, and Emily B Zimmerman. Excess deaths from covid-19 and other causes in the us, march 1, 2020, to january 2, 2021. *JAMA*, 325(17):1786–1789, 2021.

[19] Nazrul Islam, Vladimir M Shkolnikov, Rolando J Acosta, Ilya Klimkin, Ichiro Kawachi, Rafael A Irizarry, Gianfranco Alicandro, Kamlesh Khunti, Tom Yates, Dmitri A Jdanov, et al. Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. *bmj*, 373, 2021.

[20] India Office of the Registrar General. Report on medical certification of cause of death 2015, June 2017. URL `https://censusindia.gov.in/2011-Documents/mccd_Report1/MCCD_Report-2015.pdf`.

[21] India Office of the Registar General & Census Comissionner. Birth and death registration. URL `https://crsorgi.gov.in/web/index.php/auth/login`.

[22] National Commission on Population Ministry of Health and New Delhi Family Welfare, Nirman Bhawan. Population projections for india and states 2011 to 2036, July 2020. URL `https://main.mohfw.gov.in/sites/default/files/Population%20Projection%20Report%202011-2036%20-%20upload_compressed_0.pdf`.

[23] The reporters collective. URL `https://www.reporters-collective.in`.

[24] The wall of grief. URL `https://www.wallofgrief.org`.

[25] National family health survey, 2020.

[26] Rolando J Acosta and Rafael A Irizarry. Monitoring health systems by estimating excess mortality. *medRxiv*, 2020.

[27] Christopher T Leffler, Edward Yang, et al. Preliminary analysis of excess mortality in india during the covid-19 pandemic (update august 4, 2021). *medRxiv*, 2021.

[28] Lucas Rodés-Guirao Cameron Appel Charlie Giattino Esteban Ortiz-Ospina Joe Hasell Bobbie Macdonald Diana Beltekian Hannah Ritchie, Edouard Mathieu and Max Roser. Coronavirus pandemic (covid-19). *Our World in Data*, 2020. https://ourworldindata.org/coronavirus.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.22.21262432; this version posted August 25, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

District	Municipalities
Botad	Botad
Junagadh	Vanthali
Arvali	Modasa
Vadodara	Savali
Navsari	Gandevi
Bharuch	Amod
Valsad	Valsad
Valsad	Dharampur
Devbhoomi Dwarka	DwarkaNp
Devbhoomi Dwarka	Khabhalia
Devbhoomi Dwarka	JamRaval
Porbandar	Ranavav
Rajkot	Jasdan
Surendranagar	Dhangdhra
Amreli	Babra
Amreli	Lathi
Junagadh	Talala
Junagadh	Keshod
Junagadh	Manavadar
Surendranagar	Halvad
Kheda	Nadiad
Surendranagar	Thangadh
Patan	Patan
Arvali	Bayad
Panchmahal	Sahera
Anand	Anand
Anand	Karamsad
Anand	Anklav
Bharuch	Ankleshwar
Kachchh	Mundra
Morbi	Vakaner
Rajkot	Bhayavadar
Gir Somnath	Kodinar
Narmada	Rajpipala
Kachchh	Anjar
Surendranagar	Surendranagar
Rajkot	Gondal
Amreli	Amreli
Junagadh	Chorvad
Gandhinagar	Dahegam
Dahod	Dahod
Rajkot	Dhoraji
Banaskantha	Disa
Botad	Gadhda
Gandhinagar	Gandhinagar
Patan	Harij
Amreli	Jafarabad
Vadodara	Karjan
Arvali	Khedbrhma
Gandhinagar	Mansa
Sabarkantha	Prantij
Amreli	Rajula
Valsad	Umargam
Junagadh	Visadar

Supplementary Table 1: Municipalities used in the analysis.