Estimating excess deaths during the COVID-19 pandemic in Pune, India

Summary

Data from institutions across the world suggest that officially reported COVID-19 death figures may be an undercount. Establishing true mortality during crisis events like pandemics is key to improving data transparency and strengthening public health systems to effectively tackle future epidemics. This study integrated findings from three sources: statistical modeling, public surveying, and media reports to estimate excess deaths during the COVID-19 pandemic in the Pune Municipal Corporation (PMC). The results suggest an estimated **17,633 excess deaths** [95% confidence interval: 15,700 to 19,566] in the PMC region in the period from March 2020 to December 2021. An estimated **1.9 times more deaths** [95% confidence interval: 1.7 to 2.1] occurred during this time period compared to reported COVID-19 deaths.

This study was conducted by Jnana Prabodhini Foundation's Analytics wing in collaboration with the Pune Knowledge Cluster and Carnegie Mellon University.

TABLE OF CONTENTS

Background

Estimating the true mortality burden of COVID-19 is critical for establishing the true impact of the pandemic. Reports of unaccounted COVID-19 deaths have been recognized globally; conceivably due to reasons such as lack of testing, absence of medical certification, and deaths occurring beyond the purview of respective healthcare systems[1, 2]. In addition to deaths caused through active COVID-19 infection, pandemic related deaths not attributable to the virus were also noted. These deaths were attributed to delays or lack of access to healthcare, reduced hospital capacity, increased risk of diseases, and post-COVID-19 complications[3]. Together these deaths denote the pandemic associated excess deaths. In our study, we used this measure to probe whether the observed overall mortality figures were higher than what could have been expected in normal times i.e., had there been no pandemic [3, 4]. Our computational social science study integrates statistical and epidemiological modeling, wisdom of crowds public surveying, and print media reports (Fig. 1) to estimate excess deaths during the COVID-19 pandemic in the area under the administration of PMC.

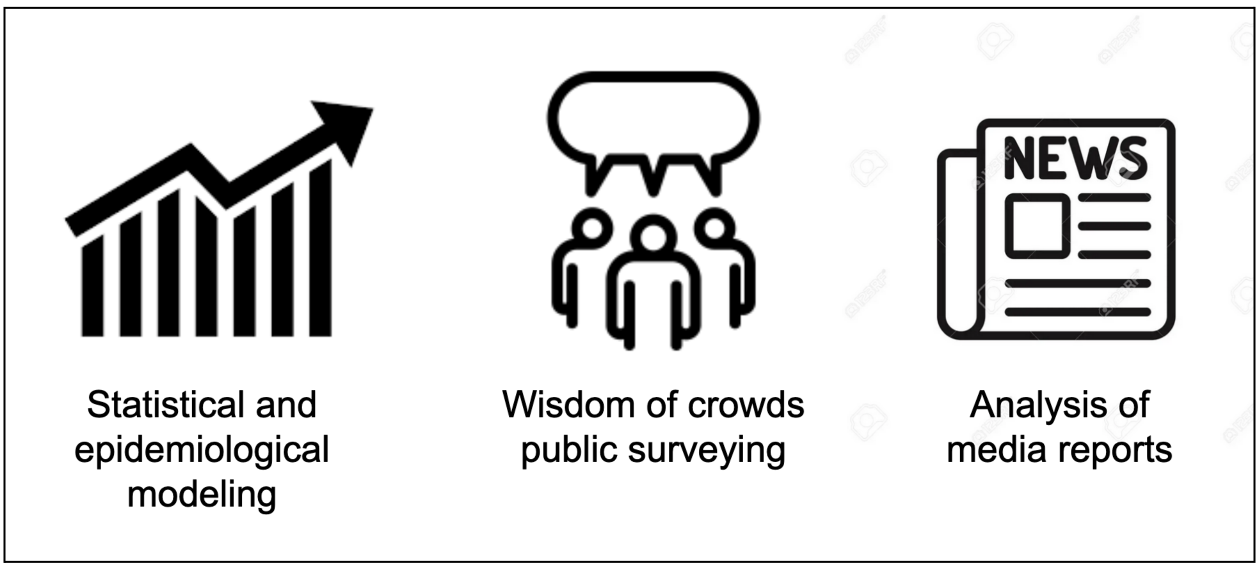

Figure 1. The excess deaths were computed by integrating three approaches: statistical and epidemiological modeling, wisdom of crowds public surveying, and analysis of print media reports.

Statistical and Epidemiological Modeling

We used an established approach to compute pandemic-associated *excess deaths*[5, 6, 7], i.e., the additional number of deaths during the COVID-19 pandemic compared to what could have been expected in pre-pandemic times. All-cause mortality figures from Jan 2014 to Dec 2021 in the Pune region were provided to us by the Pune Knowledge Cluster (PKC), a Science and Innovation Cluster set up by the Office of the Principal Scientific Advisor, Government of India[8]. PKC ultimately obtained this dataset from the Pune Municipal Corporation (PMC) Health Office's death certificate registration data. We first computed *expected deaths* by extrapolating from pre-pandemic mortality figures (January 2014 to February 2020). Further, the data from Pune's first and second COVID-19 waves (March 2020 to December 2021) were divided into COVID-19 and non-COVID-19 deaths. Following are important summary statistics from PKC's dataset from period March 2020 to December 2021:

$total\ reported\ deaths$	$reported\ COVID\ deaths$
$74,289$	$9,093$

Next, we computed *excess deaths*, which is the difference between the officially reported overall death figures and *expected deaths*. We also computed the *undercount factor*, the ratio of *excess mortality* to the officially reported COVID-19 death figures. Wherever applicable, we have included the lower and upper margins of 95% confidence intervals associated with our estimates. In the following sections, we have described two of the statistical models that we used and the key results obtained from them.

$$excess\ deaths\ =\ total\ \ reported\ deaths\ -\ \ expected\ deaths$$

$$undercount\ factor\ = \frac{excess\ deaths}{reported\ COVID\ deaths}$$

Simple Average Model

Simple average model is a simple nonparametric model that was utilized to compute the expected number of deaths for any given month during the pandemic. The expected deaths for a given month were then derived as the mean number of total deaths recorded during that month for the previous six years. The expected deaths for any month *M* from March 2020 to December 2021 is:

$$expected\ deaths\ = \frac{1}{6}\left(M_{2014} + M_{2015} + M_{2016} + M_{2017}\ + M_{2018}\ +\ M_{2019}\right)$$

where M_i is the number of deaths in month M in year i

The results of the simple average model from March 2020 to December 2021 are:

$excess\ deaths$	$excess\ deaths$ 95% $confidence\ interval$	$undercount\ factor$	$undercount\ factor$ 95% $confidence\ interval$
$19,898$	$(14,168\ to\ 25,627)$	2.2	$(1.6\ to\ 2.8)$

Overdispersed Poisson Model

In order to get closer to the real-world characteristics of the data, we implemented an overdispersed Poisson model commonly used in epidemiology and public health. This is a model that can account for variations caused by seasonal trends and population change[3, 10, 11]. The overdispersed Poisson model computed the expected deaths for any month M from March 2020 to December 2021 as:

$$expected\ deaths = P_M \cdot e^{(\alpha_M + s_M)}$$

where P_M is the population in month M, α_M is gradual trend accounting for the increasing life expectancy, and s_M is a seasonal trend accounting for seasonal variation in deaths

We obtained estimates about Pune's monthly population from the World Population Review[12]. The Overdispersed Poisson model was implemented using the *excessmort* package in R[13]. Model parameters: *harmonics* = 0.5; *knots* = 2.

The results of the overdispersed Poisson model from March 2020 to December 2021 are:

$excess\ deaths$	$excess\ deaths$ 95% $confidence\ interval$	$undercount\ factor$	$undercount\ factor$ 95% $confidence\ interval$
$14,098$	$(10,930\ to\ 17,406)$	1.6	$(1.2\ to\ 1.9)$

The results from the overdispersed model showed a marked increase in the expected death rate during the pandemic (Fig. 2).

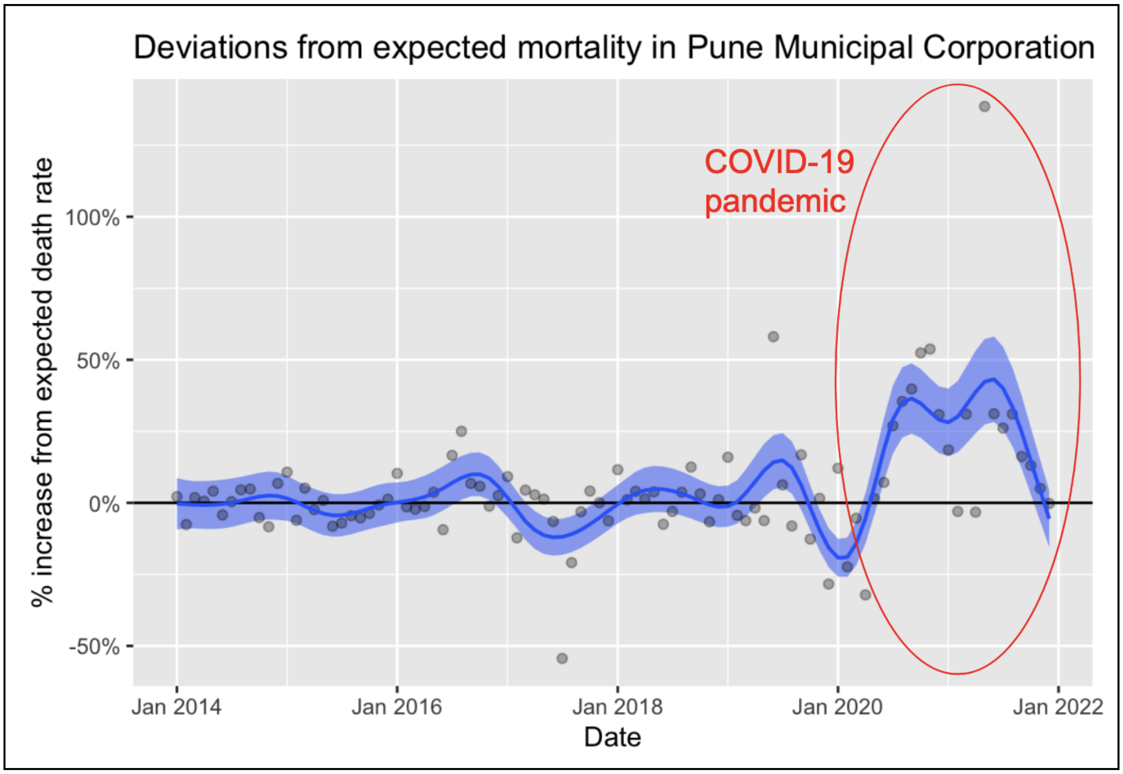

Figure 2. Deviations from the expected death rate over time: results from the overdispersed Poisson model. The blue bands represent the 95% confidence interval. The gray dots show monthly deaths relative to the expected death rate for that month.

<u>Wisdom of Crowds Public Surveying</u>

Cognitive estimation is the ability to provide reasonable answers to questions for which specific answers are not readily available[14, 15]. Wisdom of crowds, the accuracy of aggregated public opinions or collective cognitive estimates, is a phenomenon commonly studied by researchers from cognitive science, social psychology, behavioral economics, and network science[16, 17, 18, 19, 20, 21]. Recent research has used the wisdom of crowds approach in public health and epidemiology, specifically COVID-19[22, 23]. However, to our knowledge, there is no existing literature that harnesses the wisdom of crowds to extract collective cognitive estimates about excess deaths during the COVID-19 pandemic.

We addressed this knowledge gap by conducting an online wisdom of crowds survey.[24] We received ethics approval for conducting this survey from Carnegie Mellon University's Office of Research Integrity and Compliance (IRB Registration No: IRB00000352). We took care to adhere to local customs and norms. The survey was conducted in Pune from 8 January 2022 to 8 February 2022. Respondents could take the survey, which was hosted on the SurveyMonkey platform (now Momentive), in either Marathi or English. The survey was deployed in the area(s) under the jurisdiction of PMC. We employed a sample-of-convenience snowball-sampling method. We promoted the survey via social media platforms such as WhatsApp and Facebook. 280 adult respondents.

Survey respondents were asked this question: "As of January 1 2022, there have been 9117 COVID-19 deaths in Pune during the pandemic. This data is from the official government figures released by Pune Municipal Corporation (PMC). What do you think is the true number of COVID-19 deaths in Pune (as of January 1 2022)? Please choose a number between 0 and 90,000." The Marathi wording of the question was "१ जानेवारी २०२२ पर्यंत पुण्यामध्ये एकूण ९११७ कोविड-१९ मृत्यू झाले आहेत. हा सरकारी आकडा आहे (स्रोत: पुणे महानगर पालिका म्हणजेच म.न.पा). तुमच्या मते १ जानेवारी २०२२ पर्यंत पुण्यामध्ये कोविड-१९ मृत्यूंचा खरा आकडा काय असावा? ० ते ९०,००० मधील एक आकडा निवडून ते उत्तर लिहावे."

$$undercount\ factor\ = \frac{crowd\ COVID\ death\ estimate}{reported\ COVID\ deaths}$$

<u>The results of the wisdom of crowds survey are:</u>

$crowd\ death\ estimate$	$crowd\ death\ estimate$ $95\%\ confidence\ interval$	$undercount\ factor$	$undercount\ factor$ $95\%\ confidence\ interval$
$18,903$	$(17,900\ to\ 20,800)$	2.1	$(1.9\ to\ 2.3)$

The following table shows the demographics of our survey sample.

Gender	Males	151 (54%)
	Females	129 (46%)
Age	18-35	115 (41%)
	36-55	115 (41%)
	55+	50 (18%)
Survey language	Marathi	139 (50%)
	English	141 (50%)
Number of rooms in home	1-2	16 (6%)
	3	60 (22%)
	4	102 (36%)
	4+	102 (36%)
Occupation	Students	71 (25%)
	Currently employed	175 (63%)
	Healthcare workers	25 (9%)
COVID-19 +ve		101 (36%)
Sample size		280

<u>Analysis of Media Reports</u>

Estimating the true mortality burden of COVID-19 is a difficult and necessarily imprecise exercise. Though the true number of COVID-19 deaths may never be known, converging evidence from statistical modeling and public surveying suggests that the officially reported figures in Pune may be an undercount. We analyzed media reports about mismatches between officially reported figures and those observed in the context of cremations[25] and death compensation claims filed by families of the victims[26].

$$undercount\ factor\ = \frac{COVID\ death\ estimate\ from\ media\ article}{reported\ COVID\ deaths}$$

The findings from print media reports are:

	$undercount\ factor$
$field\ reports\ from\ crematoriums$ $(the\ Hindustan\ Times)$	1.9
$PMC\ death\ compensation\ claims$ $(the\ Indian\ Express)$	1.7

Aggregate Estimates

We used a simple bootstrap (random sampling with replacement; sampled 1000 times) to combine estimates obtained across sources. We computed two estimates: the number of *excess deaths* from the two statistical models and the public survey; and the *undercount factor* from the two statistical models, the public survey, and the two media reports.

The aggregate estimate across different sources is:

$excess\ deaths$	$excess\ deaths$ $95\%\ confidence\ interval$	$undercount\ factor$	$undercount\ factor$ $95\%\ confidence\ interval$
$17,663$	$(15,700\ to\ 19,566)$	1.9	$(1.7\ to\ 2.1)$

Comparison with other Indian Cities

We used estimates from other researchers computed using similar statistical modeling methods[27] to compare the undercount factor for Pune city with that from other Indian cities across the first two waves of pandemic.

	$undercount\ factor$
$Pune$	1.9
$Mumbai$	1.9
$Bengaluru$	2.1
$Chennai$	2.3

Conclusions

- An estimated **17,633 excess deaths** [95% confidence interval: 15,700 to 19,566] took place in the PMC region in the period from March 2020 to December 2021.
- **1.9 times more deaths** [95% confidence interval: 1.7 to 2.1] occurred from March 2020 to December 2021 in the PMC regions compared to reported COVID-19 deaths.

Limitations and Future Directions

- We have currently reported only confidence intervals. Future work will involve computing more robust margin of error metrics such as prediction intervals.
- We have currently aggregated estimates across sources using a simple bootstrap. This method fails to incorporate the uncertainty associated with each individual estimate. Future work will include improved model averaging techniques.
- We are currently working with other statistical models like LSTMs, SVMs, and decision trees. Future work will include results of hyperparameter tuning, training accuracy, and cross-validation accuracy from these models.
- We assume 100% death registration coverage in our statistical modeling and do not account for a potential dip in death registration during the pandemic.

Policy Recommendations

- Governments can compute excess death estimates using more granular mortality data.
- Wisdom of crowds surveys can be done in areas lacking robust public mortality data.

Contributions

This study would not have been possible without the support of numerous individuals and institutions. We thank the Analytics team at Jnana Prabodhini Foundation with over 40 volunteers from across the world for selflessly dedicating their time and efforts to this cause. We also thank the Pune Knowledge Cluster team for providing curated mortality data sourced from the Pune Municipal Corporation.

Organizations

Jnana Prabodhini Foundation, Pune Knowledge Cluster, Carnegie Mellon University

Principal Investigator

Abhishek Dedhe

Co-Investigators

Niramay Gogate, Atharv Naphade, Rohan Naphade, Dr. Pranav Pandit

Advisors

Dr. Jessica Cantlon, Dr. Aakash Chowkase, Dr. Subhash Lele, Dr. Kody Manke

Other Contributors

Vibha Bapat, Shreyas Chaudhari, Kshitij Deshmukh, Chinmay Jogdeo, Samvid Kurlekar, Manas Kshirsagar, Rohinee Joshi, Shravan Patankar

Acknowledgments

Dr. Suneeta Kulkarni, Dr. Raghunath Mashelkar, Dr. Priya Nagaraj, Jnana Prabodhini Pune

Contact

Feel free to reach out to survey@jnanaprabodhinifoundation.org with questions or comments!

Disclaimer

This is a working paper and therefore includes the preliminary conclusions from our analyses. Final results are subject to change and may be updated before submission to peer-reviewed academic platforms for publication.

References

1. Adam, D. (2022). COVID's true death toll: much higher than official records. *Nature*, *603*(7902), 562-562.
2. Knutson, V., Aleshin-Guendel, S., Karlinsky, A., Msemburi, W., & Wakefield, J. (2022). Estimating Global and Country-Specific Excess Mortality During the COVID-19 Pandemic. *arXiv preprint arXiv:2205.09081*.
3. Islam, N., Shkolnikov, V. M., Acosta, R. J., Klimkin, I., Kawachi, I., Irizarry, R. A., ... & Lacey, B. (2021). Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. *BMJ*, 373.
4. Rossen, L. M., Branum, A. M., Ahmad, F. B., Sutton, P. D., & Anderson, R. N. (2021). Notes from the field: update on excess deaths associated with the COVID-19 pandemic—United States, january 26, 2020–february 27, 2021. *Morbidity and Mortality Weekly Report, 70(15)*, 570.
5. Anand, A., Sandefur, J., & Subramanian, A. (2021). Three new estimates of India's all-cause excess mortality during the COVID-19 pandemic. *Center for Global Development*.
6. Banaji, M. (2022). Estimating potential excess mortality during the third (Omicron) wave of India's COVID-19 epidemic. *medRxiv*.
7. Centers for Disease Control and Prevention (CDC). (2022). Excess Deaths Associated with COVID-19. Accessed from: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm
8. Pune Knowledge Cluster. https://www.pkc.org.in
9. Bogam, P., Joshi, A., Nagarkar, S., Jain, D., Gupte, N., Shashidhara, L. S., ... & Mave, V. (2022). Burden of COVID-19 and case fatality rate in Pune, India: an analysis of the first and second wave of the pandemic. *IJID Regions, 2*, 74-81.
10. Acosta, R. J., & Irizarry, R. A. (2022). A flexible statistical framework for estimating excess mortality. *Epidemiology*, *33*(3), 346-353.
11. Farrington, C. P., Andrews, N. J., Beale, A. D., & Catchpole, M. A. (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. Journal of the Royal Statistical Society: Series A (Statistics in Society), 159(3), 547-563.
12. World Population Review. https://worldpopulationreview.com/world-cities/pune-population
13. Excessmort: Excess Mortality. https://CRAN.R-project.org/package=excessmort
14. Bullard, S. E., Fein, D., Gleeson, M. K., Tischer, N., Mapou, R. L., & Kaplan, E. (2004). The Biber cognitive estimation test. *Archives of clinical neuropsychology, 19(6)*, 835-846.
15. Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex,14(2)*, 294-303.
16. Da, Z., & Huang, X. (2020). Harnessing the wisdom of crowds. *Management Science, 66(5)*, 1847-1867.

17. Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature, 75(7)*, 450-451.
18. Golub, B., & Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics, 2(1),* 112-49.
19. Herzog, S. M., & Hertwig, R. (2014). Harnessing the wisdom of the inner crowd. *Trends in cognitive sciences, 18(10)*, 504-506.
20. Mannes, A. E., Larrick, R. P., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. *APA*.
21. Surowiecki, J. (2005). The wisdom of crowds. *Anchor*.
22. Sell, T. K., Warmbrod, K. L., Watson, C., Trotochaud, M., Martin, E., Ravi, S. J., ... & Servan-Schreiber, E. (2021). Using prediction polling to harness collective intelligence for disease forecasting. *BMC public health, 21(1)*, 1-9.
23. Turiel, J., Fernandez-Reyes, D., & Aste, T. (2021). Wisdom of crowds detects COVID-19 severity ahead of officially available data. *Scientific Reports, 11(1)*, 1-9.
24. Dedhe, A. M., Chowkase, A. (2021). Pandemic, Punekars, and Perceptions: Preliminary findings of a COVID-19-related Knowledge, Attitudes, Practices, and Wisdom survey. *Jnana Prabodhini Foundation.*
25. Khairnar, A. (2021). Cremation figure belies PMC count. *Hindustan Times.*
26. Express News Service. (2022). Submit genuine documents for Covid death compensation or face legal action, warns PMC. *The Indian Express*.
27. Jha, P., Deshmukh, Y., Tumbe, C., Suraweera, W., Bhowmick, A., Sharma, S., ... & Brown, P. (2022). COVID mortality in India: National survey data and health facility deaths. Science, 375(6581), 667-671.