Open Defecation is *Negatively* Associated with Reported COVID-19 Deaths in India. Are Shared Toilets the Reason?

Jason T. Kerwin, Audrey M. Dorélien, and Divya Pandey*

June 30, 2025

Abstract

Early in the COVID-19 pandemic, scientists and policymakers advocated for the reduction of open defecation to curb the disease's spread. However, SARS-CoV-2 is an airborne pathogen that can be transmitted by fecal aerosols, making poorly ventilated shared latrines a potential risk factor. Consequently, open defecation may mitigate COVID-19 transmission. Our study shows that district-level COVID-19 deaths in India are negatively, rather than positively, correlated with open defecation rates. We also show that access to private toilets is associated with reduced individual-level mortality during the pandemic, but shared toilets do not show the same protective effect. Our results suggest that as with other diseases, private toilets reduce COVID-19 mortality, but the role of shared toilets and open defecation in relation to COVID-19 specifically is more nuanced. More broadly, our results show that public health measures must be targeted to the specific disease in question: what works for one disease may fail or even backfire when applied to others.

*Kerwin: University of Washington, IZA, and J-PAL (jkerwin@uw.edu); Pandey: University of Minnesota Twin Cities; Dorélien: University of Washington. We thank Elizabeth Wrigley-Field, our discussant Aashish Gupta and the session audience at the Population Association of America 2023 conference for helpful comments. We also thank Britta Augsburg and Paul Rodriguez-Lesmes and their coauthors on Augsburg et al. 2023 for sharing the data on the WSP Total Sanitation Campaign scores and Ben Arnold and his coauthors on Patil et al. 2014 for sharing their analysis code. The authors gratefully acknowledge support from the Minnesota Population Center (P2C HD041023) funded through a grant from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD).

1 Introduction

Open defecation is typically defined as the practice of defecating in open fields, waterways, and open trenches without any proper disposal of human excreta (Saleem, Burdett, and Heaslip 2019) and is a major public health issue in South Asia. In India, rates of open defecation have drastically decreased in recent years, falling from 39% in 2015 to 19% in 2019 (PTI 2022). Despite this progress, and the Government of India's mission to make India Open Defecation-Free, hundreds of millions of people still practice open defecation in the country (Kashiwase 2023; Goff, Ahmad, and Patel 2020). While there has been a major push to construct toilets or latrines under the Government of India's *Swachh Bharat Mission*, open defecation remains a challenge to eradicate in the country due to a complex array of factors including casteism, ritual purity, and safety concerns (Coffey et al. 2014; Spears 2020).

Early in the COVID-19 pandemic, both scientists and policymakers advocated for the reduction of open defecation to reduce the spread of the disease (e.g., Sun and Han 2021). However, SARS-CoV-2 is an airborne pathogen (Nissen et al. 2020) that can be transmitted by fecal aerosols, making shared toilets a potential risk factor (Amoah et al. 2021; Huang et al. 2021; Islam et al. 2020).[1] In contrast, while open defecation increases the risk of disease transmission in general through improper handwashing, as well as fecal contamination of water bodies, it occurs outdoors and therefore, may reduce the risk of infections caused by airborne fecal aerosols or by shared exhaled air. In this chapter, my coauthors and I examine the relationship between open defecation and COVID-19 mortality in the high open defecation context of India, and explore the potential role of shared sanitation facilities in the transmission of airborne diseases.

Our main analysis examines the relationship between COVID-19 deaths and open defecation using two different mortality measures, each with its own strengths and limitations. First, we use district-level crowdsourced data on reported COVID-19 deaths in urban and rural India and find a significant negative association between reported deaths and rural open defecation. Second, we use estimates of excess mortality—a widely preferred metric

[1] Fecal aerosol transmission of coronaviruses has been well-known for some time. It was documented as a factor in the 2003 SARS-CoV-1 outbreak in Hong Kong.

to address underreporting concerns. While we observe a similar negative relationship between open defecation and excess mortality in rural areas using this measure as well, these estimates are noisier and not statistically significant. In both cases, we find no substantial evidence of a comparable relationship between open defecation and COVID-19 mortality in urban households. In our analyses, we account for demographic differences by standardizing mortality rates by age and sex. We also annualize all rates and report them as averages per 100,000 population to account for variation in the duration of reporting periods and population sizes.

Additionally, we test the hypothesis that, relative to open defecation, shared toilets pose a greater risk for COVID-19 transmission (Caruso and Freeman 2020; Hayashi et al. 2025), using individual-level mortality data from the Demographic and Health Survey (DHS) (also known as the National Family Health Survey) collected in India during the pandemic. Considering that COVID-19 mortality is exponentially higher among older people, and because the DHS data allow for age-specific analysis, we disaggregate the analysis studying the reported use of (private and shared) toilets and the probability of death by age group.[2] Our findings are noisy, but we do see that the use of private toilets at the household level is associated with lower mortality relative to open defecation across nearly all age groups in both urban and rural areas. The results for shared toilets, on the other hand, are less robust across age groups.

India does not maintain an official, structured, district-level database for COVID-19 deaths. The three data sources we use here to fill in that gap—(i) reported COVID-19 deaths from crowdsourced data, (ii) excess mortality from civil registration, and (iii) individual-level mortality from survey data during the pandemic—are imperfect sources each with their advantages and limitations. Reported COVID-19 deaths cover the entire country, but are undercounts of the actual death toll of the virus. Further, this underreporting may be correlated with observed and unobserved socio-economic factors, such as poverty. Our excess mortality estimates, on the other hand, while likely to be less correlated with poverty measures, cover only half of the country, and there are issues with civil registration data in

[2] It is not possible to conduct an excess mortality analysis using this dataset, because the district-level sample sizes are too small.

some states. The individual-level mortality data allow us to examine age gradients in the correlation between defecation practices and deaths during the pandemic, which is important given the age profile of COVID-19 mortality. However, these data cover only about 50% of districts in India and is not a representative sample: it includes only those districts that happened to be surveyed after the pandemic started. Moreover, given that the data capture deaths from all causes rather than only COVID-19 deaths, the results also capture the broader health benefits of toilet use in preventing other diseases.

Despite the limitations with data quality and availability, our findings generally suggest that there is no evidence that open defecation is a risk factor for COVID-19 mortality in India. Instead, we observe a negative correlation between the reported COVID-19 death rate and open defecation. This pattern is clearest in the crowdsourced data on reported COVID-19 deaths, where the correlation is statistically significant even with our full set of controls. It is also visible in the excess mortality data, but the correlations are insignificant because the confidence intervals are more than ten times wider. In the data on individual mortality, the picture is more nuanced—likely because that data source also captures non-COVID-19 deaths, including many diarrheal diseases for which open defecation is a serious risk factor.

This study contributes to a growing literature at the intersection of COVID-19 and Water, Sanitation, and Hygiene (WASH) in developing countries (Das et al. 2020; Desye 2021; Gwenzi 2021; Parikh et al. 2020). In a recent study, Chatterjee and Mande (2025) explore multiple state level factors associated with COVID-19 prevalence in India, and note the robust correlation between "open sanitation systems" and lower reported COVID-19 deaths. Our study, however, takes a much deeper and focused look at this relationship by using (i) district-level data; (ii) age-sex standardizing the mortality data; and (iii) exploring this association in depth using multiple data sources and mortality outcomes.

The high public health costs of open defecation are well established, and efforts to eliminate the practice remain unquestionably necessary and should continue to be prioritized. However, in the context of COVID-19 transmission in India, we find evidence that it did not exacerbate the spread of SARS-CoV-2 virus, and in fact, may have been protective compared to the use of poorly ventilated shared toilets. These findings highlight the need to consider sanitation practices in planning for future respiratory pandemics. Even if the relationships we observe are merely correlational rather than causal, our findings emphasize the need for nuanced public health messaging, as strategies effective for one disease may not be appropriate for another.

2 Data

Our main outcomes of interest are district-level COVID-19 mortality rates and excess mortality rates. For our main analysis, we use data from two different sources: (i) crowdsourced COVID-19 death reports and (ii) all-cause mortality from vital records, which we use to estimate excess mortality. These sources cover different time periods, and do not cover complete calendar years. Moreover, we are missing some months of data for certain districts. Consequently, to make our estimates more comparable, we annualize the data by calculating monthly death averages and multiplying by 12. To adjust for differences in district population size, we present deaths per 100,000 individuals per year. In all cases we divide by the district population size in 2011 (the latest Census data available). Thus, we implicitly rely on the assumption that population sizes have changed in a similar way across districts over the past 10 years. Below, we provide more details about the different sources of data and outcomes we use. Appendix Table A1 shows the availability of data for Indian states across these different outcome measures.

2.1 Annualized COVID-19 Mortality Rates

The COVID-19 mortality rate data come from publicly available data on daily district-level cases and deaths from covid19bharat.org, a data repository of pandemic-related information for India (Agarwal et al. 2021).[3] During the pandemic, state governments in India released daily health bulletins with COVID-19-related data, but these were often unstructured and not presented in a user-friendly format (Agarwal et al. 2021). The COVID19bharat dataset resulted from a large effort from volunteers who scraped daily data from state bulletins

[3] As of this writing, the website covid19bharat.org is no longer accessible. Crowdsourced data efforts first began in March 2020 through covid19india.org and then transitioned to covid19bharat.org in October 2021. The team ended the operations in January 2023 due to low COVID cases and less user engagement, among other reasons (Vivek 2023).

and other sources and made them available in a user-friendly format. While the COVID-19 figures from this dataset are an undercount (as are all counts of reported cases), it is the most comprehensive dataset on the pandemic available for India and has been used by other recent studies on COVID-19 in the country (Pandey, Gu, and Ramaswami 2022; Chatterjee and Mande 2025). We first present our analysis using data on the cumulative total number of COVID-19 deaths (half a million) reported from April 26, 2020, to May 19th, 2022 (the last day we had downloaded this dataset). The dataset covers 32 of the 36 states and union territories in India. There are 572 districts and the average district has a cumulative total of about 900 cumulative COVID-19 deaths during this period. This estimate is in line with the 525,000 COVID-19 deaths reported by the World Health Organization for India during this period (World Health Organization 2023). This is an undercount of the actual mortality burden of COVID-19 by roughly a factor of 10. For example, Deshmukh et al. (2021) estimate over 3 million COVID-19 deaths in India between June 2020 and June 2021.

2.2 Annualized Excess Mortality Rates

We construct measures of excess mortality using all-cause mortality data. The data on all-cause deaths come from the Development Data Lab (DDL)[4] which provides district-level figures on all-cause mortality for recent years, including each month in 2018, 2019, and 2020, and about six months in 2021. These data on DDL were sourced from journalists and Right to Information (RTI) activists in the country based on the states' Civil Registration Systems (CRS) that record all-cause mortality at the local level and were available for about half the country's states and 319 districts at the time we accessed the data in 2022 (Asher et al. 2020). Unlike the reported COVID-19 death data, our all-cause mortality data is thus available for only about half the country's states and districts.

Compared to reported COVID-19 deaths, excess mortality is a more accurate measure of the true impact of the COVID-19 pandemic, especially in places like India that had severe undercounts of COVID-19 cases and deaths (Leffler et al. 2022). It measures how mortality rates compare to expected mortality rates, based on historical trends for a given country or

[4] For a discussion on the different mortality data sources in India see Malani and Ramachandran (2022) and Anand (2021).

region. Similar to all-cause mortality, it also provides a better estimate of the direct and indirect impacts of COVID-19 on mortality than reported COVID-19 deaths (Leffler et al. 2022). Furthermore, excess mortality has an advantage over all-cause mortality because it can be compared across countries (Beaney et al. 2020). We calculate excess mortality for 2020 and 2021 separately by subtracting off the respective district-level mortality rate for the same calendar month, averaged across 2018 and 2019.[5] By calculating the calendar month-specific mortality rate for each district in the pre-pandemic period, this procedure also accounts for seasonal variations in mortality.

2.3 Indirect Standardization of Mortality Rates

The risk of dying from COVID-19 increases exponentially with age (Sasson 2021) and is higher for males than females (Chang et al. 2022). Therefore, age and sex are crucial potential confounders of any observed patterns in COVID-19 mortality rates. To address this, we adjust our district-level reported COVID-19 mortality and excess mortality rate figures for differences in the age and sex composition of the districts in our sample. We do not have age- or sex-specific death data for each district—only total deaths—and thus we cannot use direct standardization, which would require applying the same population composition weights to all the district-specific age-specific mortality rates. Instead, we use indirect standardization to adjust the district-level deaths to account for the differences in age and sex composition. With indirect standardization, we estimate how many deaths would occur in a district if it had the same age and sex composition as a standard population (Rodríguez 2017). This allows us to compare the actual mortality rates across districts, adjusting for differences in their demographic structures.

To carry out indirect standardization, we use the 2011 Census data to generate the age/sex composition of each district in our datasets using five-year age groups (0-4 to 85+). The standard mortality levels come from the 2018 India-wide age/sex-specific mortality rates (five-year age groups 0-4 to 85+). First, we calculate the expected mortality rates by district, assuming that each district experienced the age/sex specific mortality rates of the overall

[5] For example, March 2020 excess mortality is equal to March 2020 all-cause mortality minus average of March 2018 and March 2019 all-cause mortality.

Indian population in 2018 as follows.

$$E_d = \sum_{a,s} c_d^{a,s} m_{india}^{a,s} \tag{1}$$

In this equation, E_d represents the expected number of deaths in district d. The term $c_d^{a,s}$ denotes the population composition in district d for individuals in age group a and sex s from the 2011 census data. The variable $m_{india}^{a,s}$ refers to the national mortality rate in 2018 for the corresponding age-sex group, which serves as the standard rate. By multiplying the district's population in each age-sex category by the national mortality rate for that group and summing across all age and sex groups, we obtain the expected number of deaths in each district if it had the national mortality.

Second, we create a standardized mortality ratio (SMR) for each district, SMR_d, by taking the ratio of the actual district crude death rate, CDR_d, and dividing it by the expected mortality rate, E_d. An SMR above 1 means that the district has higher mortality than expected based on national age- and sex-specific rates. An SMR below 1 means the district has lower mortality than expected.

$$\text{SMR}_d = \frac{\text{CDR}_d}{\sum_{a,s} (c_d^{a,s} m_{india}^{a,s})} \tag{2}$$

Third, we obtain the indirect standardized rate, ISR_d, by multiplying the SMR for each district (SMR_d) by the national crude death rate in India in 2018 (CDR_{india}), which was 6.2 per 1,000.

$$\text{ISR}_d = \text{SMR}_d \times 6.2 \tag{3}$$

In summary, the equation for the indirect standardized rate (measured per 1,000), including all its components, is as follows.

$$\text{ISR}_d = \text{CDR}_{india} \frac{\text{CDR}_d}{\sum_{a,s} (c_d^{a,s} m_{india}^{a,s})} \tag{4}$$

We make some assumptions to carry out the standardization. First, we assume that the age-by-sex pattern of deaths did not change due to COVID-19. This assumption is supported by the fact that the age pattern of COVID-19 deaths follows a similar age-gradient as that of all-cause mortality (Goldstein and Lee 2020). Additionally, we assume that the age-by-sex pattern of mortality remained stable across years, as we use 2018 mortality data for standardization across all years in our datasets. We also assume that the age-by-sex composition of the districts did not significantly change between 2011 and 2022.

Finally, for the COVID-19 death data, we apply age-sex standardization to cumulative deaths and then annualize the figures. For the excess mortality data, we standardize the monthly data before annualizing it. The annualization process accounts for seasonal variations in mortality rates and ensures comparability across different outcomes, as data for various outcomes do not cover full calendar years.

2.4 Individual-level Mortality Data

In addition to the crowdsourced COVID-19 mortality data and the excess mortality data, we also construct an individual-level mortality measure indicating the death of a household member in 2020 or 2021, using data from the 2019-21 Indian DHS, also known as the National Family Health Survey (NFHS-5). The DHS collects information on deaths of household members in the household records. We leverage the fact that approximately half of India's states were surveyed in 2020 or later to identify deaths that occurred during the pandemic period. To analyze mortality during the COVID-19 pandemic, we follow the approach in Gupta et al. (2023) and restrict the analysis to deaths that took place in 2020 or later.[6] This means that we only have a subset of 17 states or union territories represented (Appendix Table A1). These data include age at death, allowing us to disaggregate our results by age group, which is not possible for our other data sources.

We also explored, and rejected, the possibility of constructing excess mortality rates using the DHS survey data. While Gupta et al. (2023) use this dataset to study excess mortality, they work with higher-level aggregations of the data. Our analysis focuses on district-level differences in open defecation rates, and so we would need to construct district-level excess mortality rates. However, the samples are so small at that level that many age categories have

[6] Our data contain three states not included in Gupta et al. (2023) Jammu & Kashmir, Andaman & Nicobar Islands, and Lakshadweep.

no deaths at all, making the construction of age-standardized mortality rates impossible.

2.5 Treatment Variable: Open Defecation Rate

For the main analysis on reported COVID-19 deaths and excess mortality, we construct our measure for open defecation, which also comes from the DHS data. We construct district-level measures of the average rate of open defecation using the DHS variable on toilet facilities used by the surveyed households. This variable is a categorical variable indicating the type of sanitation facility used by the household, including no facility or field use. We classify all households reporting no facility or field use as practicing open defecation. We define two measures of the open defecation rate for each district: one for rural parts of the district and one for urban areas. We draw this distinction because the determinants of the spread of COVID-19 are very different between rural and urban settings (Chang et al. 2022), and also because the DHS geo-identifiers distinguish between rural and urban parts of the same district.

The DHS data were collected from June 2019 to June 2021, with about 50% of the households surveyed during the pandemic period of 2020-21. Nevertheless, we use all the observations over this period in the DHS survey data to construct the treatment variables and covariates. Our analysis relies on the assumption that COVID-19 did not affect open defecation. This is likely to be the case because studies have indicated that most households did not change their defecation practices during the pandemic (Ashraf et al. 2020). The alternative would be to rely on the previous wave of the DHS (2015-16), which would increase the measurement error in our open defecation rate variable.

2.6 Control Variables

We use several district-level variables as controls in our regressions in our analysis of reported COVID deaths and excess mortality. Most of these controls come from the DHS survey, which allow for identification of both district and rural/urban status. We thus construct district-level rural and urban means for each of these controls. Key controls include the district-level fraction (separately for rural and urban areas) of households that: (i) have male householdheads; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; and (v) have a Below Poverty Line (BPL) card. We also control for the district-level means of (vi) household size; (vii) number of rooms for sleeping; and (viii) population density in 2020 separately for rural and urban areas (obtained from IPUMS DHS (Boyle, King, and Sobek 2022)).

We control for population density and the average household size in a district to account for the fact that the risk of virus transmission is higher in densely populated areas and larger households (Imdad et al. 2021). We also control for education and caste as proxies for socio-economic indicators that can in turn influence people's awareness regarding the virus, as well as their ability to take measures to reduce transmission. We do not control for the DHS wealth index since it includes measures of toilet type and quality as inputs (Croft 2017), which our treatment variable is directly related to and partially based on. Finally, we do not control for age because the outcome variable has already been standardized to account for the district-level variation in age compositions. This simplifies our analysis substantially because the relationship between age and COVID-19 mortality is non-linear and controlling for all the age-by-sex groups in our regressions would require adding numerous controls, which could lead to overfitting issues.

For the analysis of individual-level mortality, we use the same controls as those for reported COVID deaths and excess mortality, but construct them at the household level instead of using district-level means. Since the individual-level mortality analysis is done separately by age group, we do not control for age in those regressions, nor do we age-standardize mortality rates for those data.

3 Estimation Strategy

Our main analysis uses linear regressions of the outcome variable on the treatment of interest. Specifically, we estimate

$$Y_d = a + b * OD\_rural_d + c * OD\_urban_d + X_d\gamma + s_d + e_d \tag{5}$$

Here, the subscript d indexes districts, and Y_d is either a measure of district-level COVID-19 mortality rates or excess mortality. The DHS data report not only the district but also the rural/urban status of each household. Thus, we study two treatment variables: OD_rural_d is the share of rural households in a district that engage in open defecation, and OD_urban_d, the share for urban households. The error term is denoted by e_d.

For $\hat{b}$ and $\hat{c}$ to be consistent estimates of the causal effect of open defecation on mortality rates, OD_rural_d and OD_urban_d must be independent of e_d. All models include state fixed effects to account for unobserved state-level factors that may be correlated with the error term. In some of our analyses, we also control for a vector X_d of district-level covariates (described in Section 2.6). Our inferences are based on heteroskedasticity-robust standard errors, clustered by district, because that is the level at which the treatment varies.

For the analysis of individual-level mortality and shared toilets, we estimate the following separately for six age groups between ages 0 and 60+, and separately for urban and rural areas.

$$Y_i = a + b * shared\_toilet_h + c * private\_toilet_h + c * cluster\_toilet_c + X_i\gamma + e_i \qquad (6)$$

Here, Y_i denotes whether an individual i in the DHS survey died in 2020 or later. An individual's access to a shared toilet (i.e., the toilet facility used by the individual is used by at least one more household) and private toilet (i.e., the toilet facility is not used by another household) in the DHS survey data is denoted by $shared\_toilet_h$ and $private\_toilet_h$, respectively. Both shares of private and shared toilet use are calculated as percentages of the total sample. Since mortality in an area can be affected by community sanitation levels (Geruso and Spears 2018), we include a variable, $cluster\_toilet_c$, which denotes the mean use of toilets in a DHS cluster as a proxy for community sanitation levels.

4 Results

4.1 Descriptive Statistics

We observe a higher rate of open defecation in rural areas (mean = 23% of households) compared to urban areas (8%). However, even within rural areas, there is variation in the rates of open defecation (see Table 1). Figure 1 presents histograms of district-level open defecation rates for rural and urban regions. It is important to note that the reported toilet use in the DHS household data likely overestimates the actual toilet use among households and underestimates the prevalence of open defecation (Vyas et al. 2019). Despite toilet ownership or access, households or individuals may prefer open defecation, especially in rural areas. This preference is largely due to deeply rooted cultural norms, longstanding practices, and a perception that it is more natural or hygienic than using toilets (Coffey et al. 2014; Spears 2020). This preference can persist even when toilet facilities are available, particularly if the facilities are poorly constructed, lack privacy, or are poorly maintained. The actual prevalence of open defecation might not be captured in responses to household surveys due to social desirability bias (Vyas and Franz 2025). Although reported use of toilet facilities from household surveys does not completely capture actual sanitation practices in India, studies use it as an imperfect proxy for household sanitation practices (Geruso and Spears 2018; Spears 2020).

We further see that mortality varies widely across the country. The standard deviation of all-cause deaths is almost half its mean (Table 1). We confirm previous findings that COVID-19 mortality figures are indeed underreported: the mean annualized excess mortality rate is eight times greater than the annualized COVID-19 mortality rate. We find further evidence of measurement error in the COVID-19 data, as some districts report zero deaths, which is highly implausible. Specifically, there are four districts across three north-eastern Indian states in our sample that report zero COVID-19 deaths: two in Assam and one each in Arunachal Pradesh and Manipur. These are likely to be due to data reporting and/or collection errors.

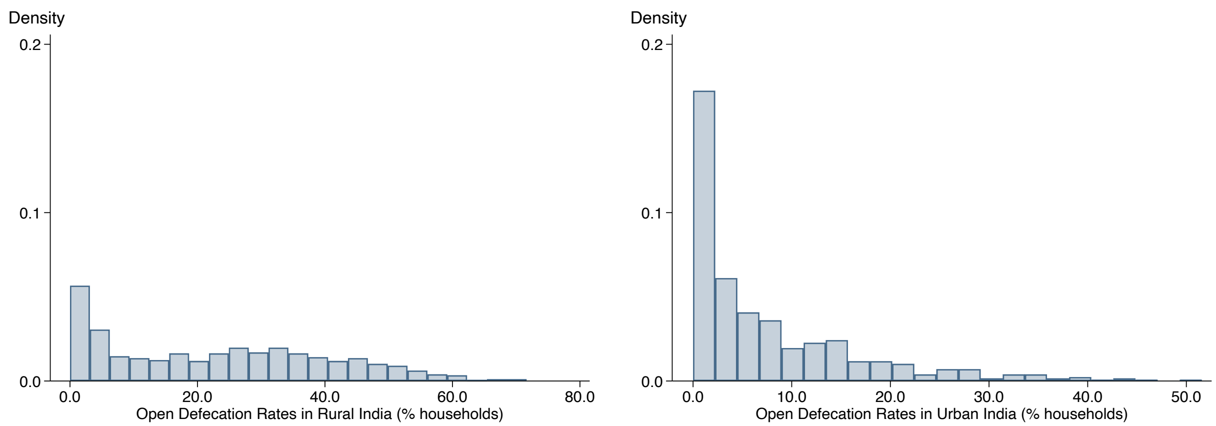
The excess mortality figures also have important reporting errors. Twelve districts in our sample report zero deaths in the underlying monthly all-cause mortality data used to construct excess mortality figures. All these districts are located in Uttar Pradesh, the country's most populous state. We also see that excess mortality is negative for 41 districts in our sample, which means that there were fewer deaths during the pandemic period than we would expect based on data from previous years. Thirty-one of these districts are in the state of Uttar Pradesh, four are in Assam, three are in Himachal Pradesh, and the rest are in Rajasthan and Maharashtra.

The pattern of all-cause deaths in our data highlights the known issues with data from Uttar Pradesh, providing an argument in favor of presenting analyses that exclude data from Uttar Pradesh. In a study of excess mortality in India during the COVID-19 pandemic, Leffler et al. (2022) note that "For Uttar Pradesh, the raw mortality data obtained from a Right-to-Information request contained anomalies, such as multiple districts with zero deaths for numerous months. Therefore, the Uttar Pradesh data were analyzed but were not included in the top-line model." Following them, we drop observations from Uttar Pradesh in our main results for excess mortality. We do not exclude Uttar Pradesh from our analyses of the other data sources, except when we are comparing them to the excess mortality results. We also test the robustness of our results to including Uttar Pradesh.

To further assess the quality of our excess mortality estimates we compare our annualized excess mortality estimates (aggregated to the state level) with annualized estimates based on Leffler et al. (2022) in Appendix Table A2. We do not expect these excess mortality estimates to be exactly the same since our estimates are based on data from March 2020 to June 2021, while their excess mortality figures are based on data from January-June 2021. Additionally, we use 2018 and 2019 as base years for the calculation of excess mortality, while Leffler et al. (2022) use 2015-2019. Despite these differences, for most states our excess mortality numbers are of a similar magnitude. Moreoever, the rankings of the states based on excess mortality rates are also generally similar. Any errors in our data are likely to be common to all data sources for India. Appendix Table A3 shows another version of this comparison of excess mortality figures with Leffler et al. (2022), where we restrict our data to January-June 2021. We see a jump in our estimates for excess mortality for this period since this was the period of the deadly Delta wave. Again, the overall rankings of excess mortality are similar to Leffler et al. (2022).

Figure 1

District-Level Open Defecation Rates in India



Panel A: Rural **Panel B:** Urban

Notes: Authors' calculations based on India DHS 2019 data.

Table 1
District Level Summary Statistics

	N	Mean	SD	Min	Max
Open Defecation (% households)					
Rural	567	23.06	17.63	0	71.64
Urban	569	7.62	9.18	0	51.57
Use of Shared Toilets (% households)					
Rural	567	7.05	5.17	0	52.38
Urban	569	10.12	6.94	0	47.79
Reported COVID-19 Deaths (April-2020 to May-2022)					
Annualized Deaths per Year	572	408	1,033	0	12,092
Age-sex Standardized	572	362	968	0	13,347
Annualized Death Rate per 100,000	572	17	24	0	293
Age-sex Standardized	572	16	20	0	243
All-Cause Deaths (March-2020 to June-2021)					
Annualized Deaths per Year	319	14,125	13,962	170	83,414
Age-sex Standardized	319	13,684	13,141	133	83,598
Annualized Death Rate per 100,000	319	579	287	54	2,035
Age-sex Standardized	319	563	268	54	2,099
Excess Deaths (March-2020 to June-2021)					
Annualized Deaths per Year	319	3,534	6,037	-18,219	42,052
Age-sex Standardized	319	3,418	5,850	-18,668	37,690
Annualized Death Rate per 100,000	319	133	156	-455	868
Age-sex Standardized	319	129	155	-469	874

Notes: Means for urban and rural open defecation are similar across both our samples for reported COVID deaths and excess mortality analysis. N is the number of districts. Deaths are reported as counts at the district level.

4.2 Effects on Reported COVID-19 Mortality

We first show results for our analysis on reported COVID-19 deaths obtained from the crowd-sourced data and find that open defecation is negatively correlated with reported COVID-19 mortality. Based on the simplest regression model presented in Table 2, an increase of 10 percentage points in open defecation rate among rural households is associated with a reduction of 5.2 COVID-19 deaths per 100,000 people over the span of 12 months (Table 2, column 1). Accounting for variation in the age and sex composition of the districts reduces the magnitude of the effect of rural open defecation on COVID-19 deaths by about 0.02 deaths. In the preferred specification in column 6, controlling for the household covariates and including state fixed effects, each 10-percentage point increase in the open defecation rate among rural households is associated with a decrease of 1.3 reported COVID-19 deaths per 100,000 people. These effect sizes are modest: In a district of one million people, this effect translates to approximately 13 fewer reported deaths. These results suggest that there is an inverse relationship between the rural household open defecation rate and reported COVID-19 deaths. Figure 2, which presents results from columns 1, 4, 5, and 6 of the table, also clearly depicts this inverse relationship. Although our analysis is based on district-level aggregated data, the results may still capture the negative spillovers of open defecation that occur at the local (neighborhood level), as fecal contamination can affect the broader community beyond individual households (Geruso and Spears 2018).

The relationship between higher rates of open defecation and the number of reported COVID-19 deaths differs between rural and urban households, as depicted in Figure 2. In the baseline model in Table 2 (column 1), an increase of 10 percentage points in open defecation rate among urban households is linked with an increase of two reported COVID-19 deaths per 100,000 people over 12 months. However, after controlling for household covariates and adding state fixed effects in the age-standardized model, we see that a 10 percentage-point increase in the rate of open defecation among urban households is associated with an increase of 0.6 deaths per 100,000 people and is not statistically significant (column 6).

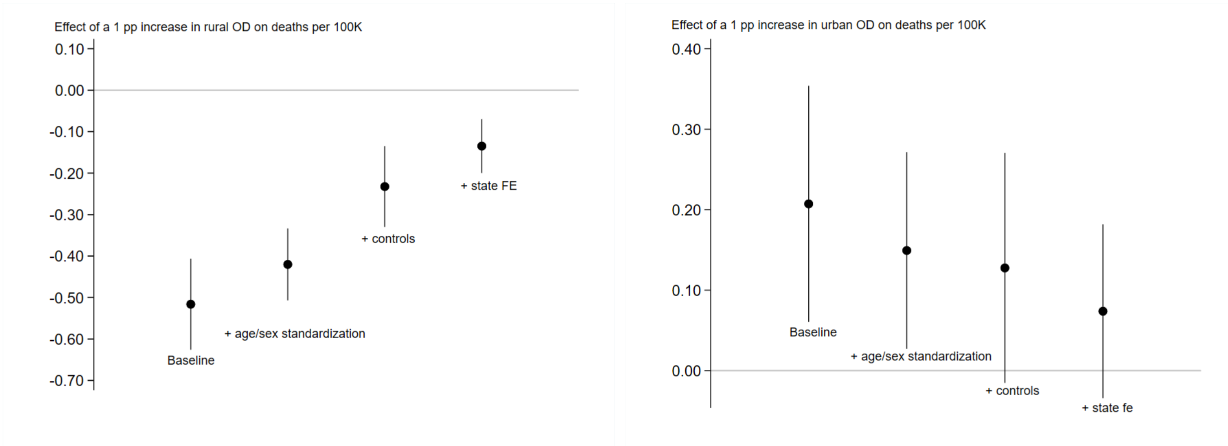
Table 2

Open Defecation and Annualized COVID-19 Mortality in India

	COVID-19 Deaths per 100K					
				Age-sex standardized		
	(1)	(2)	(3)	(4)	(5)	(6)
Open Defecation (% households)						
Rural	-0.516***	-0.260***	-0.159***	-0.420***	-0.234***	-0.134***
	(0.056)	(0.059)	(0.038)	(0.044)	(0.050)	(0.033)
Urban	0.207***	0.147*	0.040	0.149**	0.122*	0.064
	(0.075)	(0.088)	(0.060)	(0.062)	(0.073)	(0.055)
Household covariates	No	Yes	Yes	No	Yes	Yes
State FE	No	No	Yes	No	No	Yes
Observations	564	564	558	564	564	558
Adjusted R-squared	0.151	0.379	0.769	0.160	0.354	0.702
Mean of outcome variable	17	17	17	16	16	16
SD of outcome variable	24	24	24	20	20	20
Mean of Rural open defecation	23.06	23.06	23.06	23.06	23.06	23.06
SD of Rural open defecation	17.63	17.63	17.63	17.63	17.63	17.63
Mean of Urban open defecation	7.62	7.62	7.62	7.62	7.62	7.62
SD of Urban open defecation	9.18	9.18	9.18	9.18	9.18	9.18

Notes: Data on reported COVID-19 deaths are from crowd-sourced data (Agarwal et al. 2021). Controls are the district-level fraction (separately for rural and urban areas) of households that: (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020. Heteroskedasticity-robust standard errors, clustered by district, in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 2
Open Defecation and Annualized COVID-19 Mortality in India



Panel A: Rural **Panel B:** Urban

Notes: Data on reported COVID-19 deaths are from crowd-sourced data (Agarwal et al. 2021). OD refers to open defecation. Whiskers indicate 95% confidence intervals. Controls used are the district-level fraction (separately for rural and urban areas) of households that: (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020.

4.3 Excess Mortality

We summarize the results for excess mortality in Figure 3, which corresponds to columns 1, 4, 5, and 6 of the results detailed in Table 3.[7] Across all the different specifications, there is an inverse relationship between open defecation among rural households and excess mortality, but none of the estimates are statistically significant. If we take the results with no regression or age adjustments literally, an increase of 10 percentage points in the open defecation rate among rural households would imply a reduction of 5.5 excess deaths per 100,000 people over the span of 12 months. These results are very similar in magnitude to our baseline model findings for COVID-19 deaths. After controlling for the household covariates and including state fixed effects, each 10-percentage point increase in the open

[7] We also present estimates for the underlying data of all-cause mortality in Table A4. These mortality figures are caused by many other causes of death and thus cannot be attributed to COVID-19, and thus, excess mortality remains the preferred measure.

defecation rate among rural households is associated with a decrease of 9.5 excess deaths per 100,000 people per year, which is substantially larger than the corresponding figure for reported COVID-19 deaths. Although not statistically significant, we find a stronger inverse relationship between the rate of open defecation among rural households and excess death, which is thought to be a more accurate estimate of COVID-19 mortality.

The magnitude and the direction of the relationship between higher rates of open defecation in urban households and the number of excess deaths changes across the different model specifications. In the baseline model in Table 3 (column 1), an increase of 10 percentage points in the open defecation rate among urban households is linked with an increase of 14.2 excess deaths per 100,000 people over 12 months. In the age-standardized model in column 6, controlling for household covariates and adding the state fixed effects, we see that a 10 percentage point increase in the rate of open defecation among urban households is associated with an increase of just 0.7 excess deaths per 100,000 people and is not statistically significant.

The results in Figure 3 exclude data from Uttar Pradesh, but we also present results for excess mortality with Uttar Pradesh included in Appendix Figure A1. For rural areas, the results are qualitatively unchanged if we include the state. For urban areas, however, when we include data from Uttar Pradesh, the models suggest a decrease in excess deaths, but with wide confidence intervals.

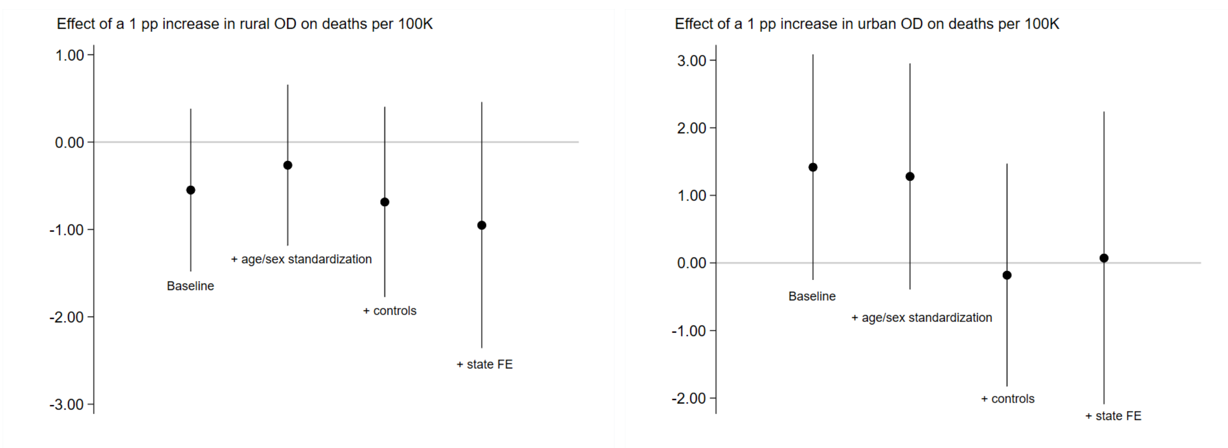
Table 3
Open Defecation Rate and Annualized Excess Mortality Rate

	Excess All-Cause Deaths per 100K			Age-sex standardized		
	(1)	(2)	(3)	(4)	(5)	(6)
Open Defecation (% households)						
Rural	-0.549	-0.773	-1.065	-0.265	-0.686	-0.952
	(0.471)	(0.552)	(0.716)	(0.466)	(0.551)	(0.713)
Urban	1.416*	-0.528	-0.014	1.279	-0.181	0.072
	(0.845)	(0.862)	(1.118)	(0.846)	(0.836)	(1.097)
Household covariates	No	Yes	Yes	No	Yes	Yes
State FE	No	No	Yes	No	No	Yes
Observations	241	241	240	241	241	240
Adjusted R-squared	-0.002	0.318	0.402	-0.003	0.299	0.384
Mean of outcome variable	160	160	160	156	156	156
SD of outcome variable	120	120	120	117	117	117
Mean of Rural open defecation	24.30	24.30	24.30	24.30	24.30	24.30
SD of Rural open defecation	17.45	17.45	17.45	17.45	17.45	17.45
Mean of Urban open defecation	6.95	6.95	6.95	6.95	6.95	6.95
SD of Urban open defecation	7.78	7.78	7.78	7.78	7.78	7.78

Notes: Data from excess mortality are from Development Data Lab (Asher et al. 2020). Controls are the district-level fraction (separately for rural and urban areas) of households that: (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020. Uttar Pradesh districts have been dropped from the analysis. Heteroskedasticity-robust standard errors, clustered by district, in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 3

Open Defecation Rate and Annualized Excess Mortality Rate in India



Panel A: Rural **Panel B:** Urban

Notes: Data from excess mortality are from Development Data Lab (Asher et al. 2020). OD refers to open defecation. Whiskers indicate 95% confidence interval. Controls used are the district-level fraction (separately for rural and urban areas) of households that: (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020. Districts from Uttar Pradesh have been dropped.

4.4 Shared and Private Toilets

We next investigate how the reported use of shared toilets, relative to open defecation, affected COVID-19 mortality. Shared toilet use, as captured by the DHS surveys, refers to the use of the toilet facility by another household. Shared toilets are typically communal sanitation facilities used by multiple households, most commonly found in poor urban areas, peri-urban areas, and some rural communities. There is a wide variation in the structure and quality of shared facilities, but they generally consist of basic concrete or brick enclosures, at times housing multiple stalls (see for example, Appendix Figure A2) and are often separated by gender. Shared toilet use can be inconsistent, influenced by factors such as seasonality (e.g. higher usage during rainy seasons), the condition and maintenance of community toilets, distance to the facilities, and wait times (Heijnen et al. 2014; Heijnen et al. 2015; Av et al. 2017). The maintenance and cleaning of shared toilets is handled through a mix of government, private, and community arrangements, depending on the facility type. For instance, in government-built facilities, municipal-level bodies are responsible for maintenance, while in urban and peri-urban areas, some shared toilets can be funded by small user fees.

Aside from issues of poor maintenance, the shared community latrines in India are generally pit latrines and are not always adequately ventilated (see UNICEF 2022). This implies that airborne viruses like SARS-CoV-2 can be transmitted through the stagnant air pocket in poorly ventilated spaces. Poor ventilation can be a common feature in improved latrines as well. For example, data on India's 1999-2012 Total Sanitation Campaign suggests that 20.2% of latrines considered "improved" did not have modern ventilation or a flush system to control sewer gas (Patil et al. 2014). The use of such shared toilets in India is not very high: 7% of the households in rural India report use of shared toilets compared to 10% for urban India (Table 1). Nevertheless, the results we see for COVID-19 deaths and excess mortality might be confounded by these other sanitation practices.

To assess how individual mortality varies with the reported use of private and shared toilets compared with open defecation, we use individual-level mortality data from the DHS. COVID-19 mortality is known to have a strong age-related pattern, with the mortality rate increasing exponentially with age (Sasson 2021). Since we do have people's ages in the DHS individual-level data (unlike the reported COVID-19 deaths and excess mortality data), we analyze individual mortality separately by age group.

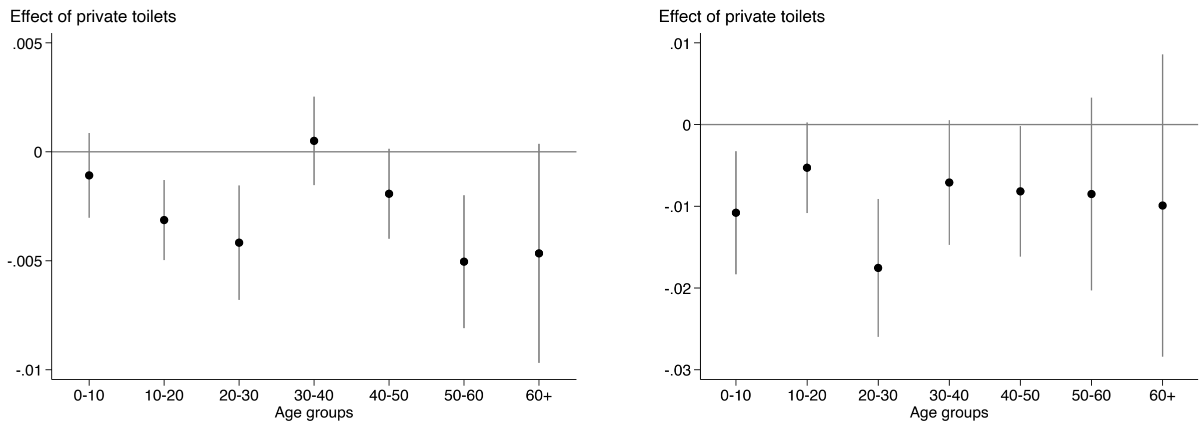
Our first finding is that the household use of private toilets compared to open defecation is negatively associated with the probability of death across most age groups, in both rural and urban areas (Figure 4). On the other hand, the reported use of shared toilets, when compared to open defecation, is negatively associated with the probability of death for children younger than 10 in both rural and urban areas, although the findings are not statistically significant (panels A and B in Figure 5). The general protective pattern of shared toilets, relative to open defecation, for the youngest age group is expected, given that the beneficial effects of sanitation on children's life expectancy in India are established (Geruso and Spears 2018). For the older age groups, however, the relationship between access to shared toilets and the probability of dying is noisy and not consistent across age groups for both urban and rural areas.

The protective effect of private toilets, with respect to open defecation, on deaths during the pandemic period emerges clearly in our analysis.[8] But we have less conclusive findings for the effect of shared toilet use on mortality. The use of shared toilets has been associated with an increased risk of adverse health outcomes compared to individual household toilets (Heijnen et al. 2014), but accurately understanding and characterizing their usage remains a challenge. Along with these factors, our results are also likely to be influenced by the quality of toilets that households have access to, specifically ventilation. Not all improved shared toilets, as defined by WHO/UNICEF Joint Monitoring Programme, are required to have ventilation, and at least 20% do not (Patil et al. 2014). The presence (or lack) of ventilation in shared toilets that individuals surveyed in the DHS had access to is likely to affect our results.

[8] This effect persists despite the potential for within-household transmission of the virus among members.

Figure 4

Reported use of Private Toilets and the Probability of Death (Relative to OD) in the 5$^{\text{th}}$ round of DHS survey in India

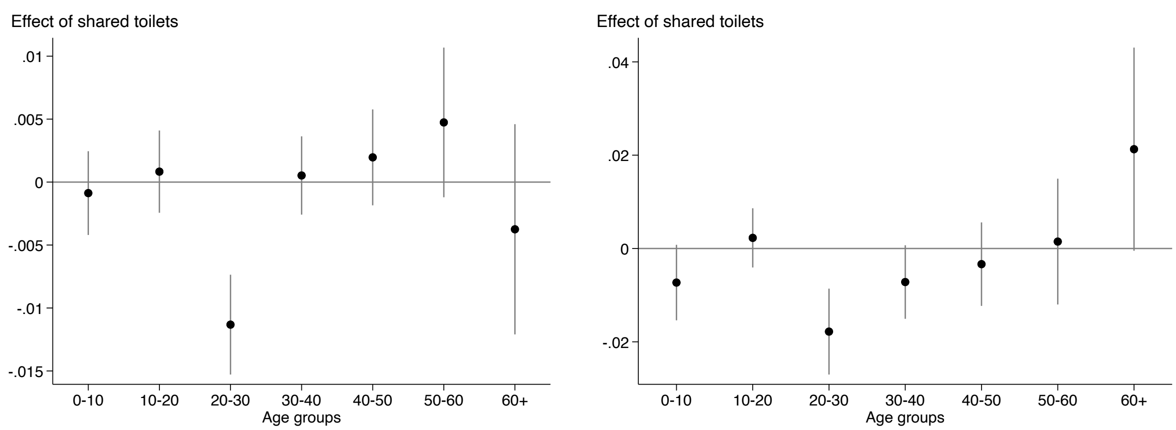


Panel A: Rural **Panel B:** Urban

Notes: OD refers to open defecation. Whiskers indicate 95% confidence interval. Controls used are the household level controls indicate whether they are (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020.

Figure 5

Reported use of Shared Toilets and the Probability of Death (Relative to OD) in the 5$^{\text{th}}$ round of DHS survey in India



Panel A: Rural **Panel B:** Urban

Notes: OD refers to open defecation. Whiskers indicate 95% confidence interval. Controls used are the household level controls indicate whether they are (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020.

5 Discussion and Conclusion

Our findings collectively provide suggestive evidence that district-level open defecation rates are negatively associated with COVID-19 deaths in rural areas of India. This association is strongest and most robust for reported COVID-19 deaths, even after we control for state fixed effects and our full set of controls. The excess mortality results for rural areas, on the other hand, also show a negative association but are not statistically significant.

One possible explanation for the difference in significance is that for the excess mortality samples, we have data from half as many districts as the reported COVID-19 deaths, so we might be limited in power for our excess mortality analysis. To test whether it is the selection of the districts that is driving the difference, as a sensitivity test, we also restrict the reported COVID-19 data to the districts included in the excess mortality data and still find statistically significant negative associations between open defecation among rural households and COVID-19 mortality with this smaller sample (Appendix Table A5). Even with this smaller sample, the standard errors are considerably smaller than those for the other mortality measures—about 0.2% of a standard deviation, as opposed to 4% for excess mortality. This means that we have much higher statistical power for the reported death measures than for any other outcome.

This difference in statistical power points to a potential advantage of using reported COVID-19 deaths instead of excess mortality: while they are biased, they may potentially be more precise, including less noise that is due to other drivers of deaths. In other words, people die for many causes, so even during the pandemic, the vast majority of deaths were not due to COVID-19. The greater precision of reported COVID-19 deaths may also be driven by the way in which they were recorded. Although COVID-19 deaths are underreported, the use of crowdsourced data may provide more consistent estimates across districts, especially in areas where official reporting is limited. However, this consistency depends on the extent of public participation and local reporting efforts. In contrast, the all-cause mortality data based on the Civil Registration System, were retrieved through separate efforts by activists and journalists. These retrieval efforts may introduce biases, especially if they focused on specific districts or lacked access to accurate civil registration records in others. As a result, there may be systematic differences in the reporting of all-cause deaths across districts. Our estimation strategy may not be able to control for important district-level differences that might influence both mortality and reporting. Another potential explanation for the precision could emerge from the differences in the time periods covered by the COVID-19 mortality and excess mortality data.

The relationship between open defecation and mortality among urban households is less consistent. Based on our most conservative model the relationship is never statistically significant—the effect size is small and positive for COVID-19 mortality and excess mortality. Our inconsistent results for urban India might be capturing the fact that open defecation practices differ between urban settings and rural areas. In rural areas, open defecation is often in the fields and a consistent norm. In urban areas, on the other hand, open defecation might not be as consistent and take place along with the use of community toilets or neighbors' toilets, and may be seasonal. Our results from individual-level data suggest that compared to open defecation, private toilets have a strong protective effect for COVID-19 as with other diseases. The relationship between individual-level mortality and shared toilets, however, is less clear. We posit that factors such as age and urban/rural status would influence the results, along with other factors such as the quality of the ventilation in the shared toilets, which we cannot capture in our analysis.

Our specifications suggest that a 10-percentage point increase in rural open defecation is associated with 1.3 to 9.5 fewer deaths per 100,000 people, depending on whether we use reported COVID-19 deaths or excess mortality as the outcome. These estimates are modest. For instance, amongst the many public health benefits of improved sanitation, studies have estimated that improved sanitation can reduce under-five or infant mortality by as much as 419 to 660 per 100,000 live births (Headey and Palloni 2019; Chakrabarti et al. 2024). This contrast is expected given the differences in disease transmission pathways, but it nonetheless underscores the potential for sanitation infrastructure to influence health outcomes beyond traditional channels.

5.1 Limitations

Our study's foremost limitation is the lack of updated, complete, or reliable data across the different aspects of our mortality analysis, which may affect our findings. First, neither of our main mortality measures include information on the ages or sexes of the people who died, which meant that we could not control for variation in age and sex using direct standardization methods. Second, the reported COVID-19 mortality data that we use were not based on data from the Civil Registration System or other official sources. Third, our excess mortality analyses use data from less than half of India's states and from a fraction of districts within the states. Finally, we use data from 2018 and 2019 to create the baseline for our excess mortality estimates. It might be more beneficial to use a longer time period to create the baseline, for instance using data from the five years prior to the start of the pandemic. Our approach also does not take into account potential time trends in mortality (Leffler et al. 2022).

We explored using life expectancy to study the relationship between sanitation practices and mortality, but it is not possible to accurately estimate district level-life expectancy using data from the DHS. We would need to rely on household survey data and a limited number of years (2020-21), and therefore for many age groups we would not be able to observe any deaths. Aggregating the data at the state level would mitigate this sampling error problem (Gupta and Sudharsanan 2022), but then we would lose the variation in our data that we actually use to estimate the open defecation-mortality relationship (within-state variation in open defecation rates).

This study emphasizes the need for continued efforts to better understand sanitation practices in India, particularly the use of shared community toilets. Moreover, the limitations in our COVID-19 mortality data underscore the crucial role of official data on mortality during pandemics. Despite data limitations, our findings highlight the need for caution in public health messaging and the importance of understanding the broader role of infrastructural investments in sanitation in shaping vulnerability to emerging health threats. For example, local efforts to limit the spread of COVID-19 in a future wave by encouraging toilet use should also highlight the transmission risk that might come with the use of shared toilets, especially for the older or more vulnerable population. Alongside more nuanced public health messaging, continued public investments in the cleanliness, ventilation, and maintenance of shared sanitation facilities should be made to mitigate the risks of future pandemics.

References

Agarwal, Mayank, Tathagata Chakraborti, Sachin Grover, and Arunima Chaudhary. 2021. "COVID-19 India Dataset: Parsing COVID-19 Data in Daily Health Bulletins from States in India." arXiv:2110.02311.

Amoah, Isaac Dennis, Leanne Pillay, Nashia Deepnarian, Oluyemi Awolusi, Kriveshin Pillay, Preshod Ramlal, Sheena Kumari, and Faizal Bux. 2021. "Detection of SARS-CoV-2 RNA Contact Surfaces within Shared Sanitation Facilities." *International Journal of Hygiene and Environmental Health*, 236: 113807.

Anand, Abhishek. 2021. "Three New Estimates of India's All-Cause Excess Mortality During the COVID-19 Pandemic." Working Paper 589, Center for Global Development.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2020. "The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG)." World Bank Economic Review (Revise and Resubmit).

Ashraf, Sania, Jinyi Kuang, Upasak Das, and Cristina Bicchieri. 2020. "Sanitation Practices during Early Phases of COVID-19 Lockdown in Peri-Urban Communities in Tamil Nadu, India." *The American Journal of Tropical Medicine and Hygiene*, 103(5): 2012–2018.

Av, Surya, Archna Vyas, Madhu Krishna, and Naseem Abidi. 2017. "Identifying Determinants of Toilet Usage by Poor in Urban India." *Procedia Computer Science*, 122: 634–641.

Beaney, Thomas, Jonathan M Clarke, Vageesh Jain, Amelia Kataria Golestaneh, Gemma Lyons, David Salman, and Azeem Majeed. 2020. "Excess Mortality: The Gold Standard in Measuring the Impact of COVID-19 Worldwide?" *Journal of the Royal Society of Medicine*, 113(9): 329–334.

Boyle, E. H., M. King, and M. Sobek. 2022. "IPUMS-Demographic and Health Surveys: Version 9 [dataset]."

Caruso, Bethany A, and Matthew C Freeman. 2020. "Shared Sanitation and the Spread of COVID-19: Risks and Next Steps." *The Lancet Planetary Health*, 4(5): e173.

Chakrabarti, Suman, Soyra Gune, Tim A. Bruckner, Julie Strominger, and Parvati Singh. 2024. "Toilet construction under the Swachh Bharat Mission and infant mortality in India." *Scientific Reports*, 14(1): 20340. Publisher: Nature Publishing Group.

Chang, Dianna, Xin Chang, Yu He, and Kelvin Jui Keng Tan. 2022. "The Determinants of COVID-19 Morbidity and Mortality Across Countries." *Scientific Reports*, 12(1): 5888.

Chatterjee, Bithika, and Shekhar C. Mande. 2025. "Demography, sanitation and previous disease prevalence associate with COVID-19 deaths across Indian States." *Scientific Reports*, 15(1): 10270. Publisher: Nature Publishing Group.

Coffey, Diane, Aashish Gupta, Payal Hathi, Nidhi Khurana, Dean Spears, Nikhil Srivastav, and Sangita Vyas. 2014. "Revealed Preference for Open Defecation." *Economic & Political Weekly*, 49(38): 43.

Croft, Trevor. 2017. "Wealth Index Calculation."

Das, A., S. Ghosh, K. Das, I. Dutta, T. Basu, and M. Das. 2020. "Re: (In)Visible Impact of Inadequate WaSH Provision on COVID-19 Incidences Cannot Be Ignored in Large and Megacities of India." *Public Health*, 185: 34–36.

Deshmukh, Yashwant, Wilson Suraweera, Chinmay Tumbe, Aditi Bhowmick, Sankalp Sharma, Paul Novosad, Sze Hang Fu, Leslie Newcombe, Hellen Gelband, Patrick Brown, and Prabhat Jha. 2021. "Excess Mortality in India From June 2020 to June 2021 during the COVID Pandemic: Death Registration, Health Facility Deaths, and Survey Data." *Medrxiv*, 2021-07.

Desye, Belay. 2021. "COVID-19 Pandemic and Water, Sanitation, and Hygiene: Impacts, Challenges, and Mitigation Strategies." *Environmental Health Insights*, 15.

Geruso, Michael, and Dean Spears. 2018. "Neighborhood Sanitation and Infant Mortality." *American Economic Journal: Applied Economics*, 10(2): 125–162.

Goff, Allyson, Akram Ahmad, and Isha Patel. 2020. "Fecal-Oral Transmission of COVID-19 in India." *Journal of Pharmacy Practice and Community Medicine*, 6(2).

Goldstein, Joshua R., and Ronald D. Lee. 2020. "Demographic Perspectives on the Mortality of COVID-19 and Other Epidemics." *Proceedings of the National Academy of Sciences of the United States of America*, 117(36): 22035–22041.

Gupta, Aashish, and Nikkil Sudharsanan. 2022. "Large and Persistent Life Expectancy Disparities between India's Social Groups." *Population and Development Review*, 48(3): 863–882.

Gupta, Aashish, Payal Hathi, Murad Banaji, Prankur Gupta, Ridhi Kashyap, Vipul Kumar Paikra, Kanika Sharma, Anmol Somanchi, Nikkil Sudharsanan, and Sangita Vyas. 2023. "Large and Unequal Life Expectancy Declines Associated with the COVID-19 Pandemic in India in 2020." SocArXiv preprint.

Gwenzi, Willis. 2021. "Leaving No Stone Unturned in Light of the COVID-19 Faecal-Oral Hypothesis? A Water, Sanitation and Hygiene (WASH) Perspective Targeting Low-Income Countries." *Science of The Total Environment*, 753: 141751.

Hayashi, Michael A. L., Sophia M. Simon, Kaiyue Zou, Hannah Van Wyk, Mondal Hasan Zahid, Joseph N. S. Eisenberg, and Matthew C. Freeman. 2025. "Shared Sanitation Facilities and Risk of Respiratory Virus Transmission in Resource-Poor Settings: A COVID-19 Modeling Case Study." *Risk Analysis*, 45(3): 638–652. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.17633.

Headey, Derek, and Giordano Palloni. 2019. "Water, Sanitation, and Child Health: Evidence From Subnational Panel Data in 59 Countries." *Demography*, 56(2): 729–752.

Heijnen, Marieke, Oliver Cumming, Rachel Peletz, Gabrielle Ka-Seen Chan, Joe Brown, Kelly Baker, and Thomas Clasen. 2014. "Shared Sanitation versus Individual Household Latrines: A Systematic Review of Health Outcomes." *PLOS ONE*, 9(4): e93300. Publisher: Public Library of Science.

Heijnen, Marieke, Parimita Routray, Belen Torondel, and Thomas Clasen. 2015. "Shared Sanitation versus Individual Household Latrines in Urban Slums: A Cross-Sectional Study in Orissa, India." *The American Journal of Tropical Medicine and Hygiene*, 93(2): 263–268.

Huang, Jianxiang, Phil Jones, Anqi Zhang, Shan Shan Hou, Jian Hang, and John D. Spengler. 2021. "Outdoor Airborne Transmission of Coronavirus Among Apartments in High-Density Cities." *Frontiers in Built Environment*, 7. Publisher: Frontiers.

Imdad, Kashif, Mehebub Sahana, Md Juel Rana, Ismail Haque, Priyank Pravin Patel, and Malay Pramanik. 2021. "A district-level susceptibility and vulnerability assessment of the COVID-19 pandemic's footprint in India." *Spatial and Spatio-temporal Epidemiology*, 36: 100390.

Islam, M. Saiful, Kazi M. Rahman, Yanni Sun, Mohammed O. Qureshi, Ikram Abdi, Abrar A. Chughtai, and Holly Seale. 2020. "Current Knowledge of COVID-19 and Infection Prevention and Control Strategies in Healthcare Settings: A Global Analysis." *Infection Control and Hospital Epidemiology*, 41(10): 1–11.

Kashiwase, Haruna. 2023. "World Toilet Day: 420 Million People Are Defecating Outdoors."

Kotwal, Nandita, Rebecca J. Gilsdorf, Nishtha Mehta, and Martin Gambrill. 2020. "To Share or Not to Share Sanitation Facilities: A Coronavirus Conundrum." World Bank Blogs.

Leffler, Christopher T., Joseph D. Lykins V, Saurav Das, Edward Yang, and Sneha Konda. 2022. "Preliminary Analysis of Excess Mortality in India During the COVID-19 Pandemic." *The American Journal of Tropical Medicine and Hygiene*, 106(5): 1507–1510.

Malani, Anup, and Sabareesh Ramachandran. 2022. "Using Household Rosters from Survey Data to Estimate All-Cause Excess Death Rates During the COVID Pandemic in India." *Journal of Development Economics*, 159(102988).

Nissen, Karolina, Janina Krambrich, Dario Akaberi, Tove Hoffman, Jiaxin Ling, Åke Lundkvist, Lennart Svensson, and Erik Salaneck. 2020. "Long-distance Airborne Dispersal of SARS-CoV-2 in COVID-19 Wards." *Scientific Reports*, 10(1): 19589.

Pandey, Bhartendu, Jianyu Gu, and Anu Ramaswami. 2022. "Characterizing COVID-19 waves in urban and rural districts of India." *npj Urban Sustainability*, 2(1): 1–9. Number: 1 Publisher: Nature Publishing Group.

Parikh, Priti, Loan Diep, Jaideep Gupte, and Monica Lakhanpaul. 2020. "COVID-19 Challenges and WASH in Informal Settlements: Integrated Action Supported by the Sustainable Development Goals." *Cities*, 107(2020): 102871.

Patil, Sumeet R., Benjamin F. Arnold, Alicia L. Salvatore, Bertha Briceno, Sandipan Ganguly, John M. Colford Jr, and Paul J. Gertler. 2014. "The Effect of India's Total Sanitation Campaign on Defecation Behaviors and Child Health in Rural Madhya Pradesh: A Cluster Randomized Controlled Trial." *PLOS Medicine*, 11(8): e1001709.

PTI. 2022. "Nineteen Percent Households Do Not Use Any Toilet Facility: NFHS Survey." *Indian Express*.

Rodríguez, Germán. 2017. "Direct and Indirect Standardization." *https://grodri.github.io/demography/Standardization.pdf*, Accessed June 13, 2025.

Saleem, Mahrukh, Teresa Burdett, and Vanessa Heaslip. 2019. "Health and Social Impacts of Open Defecation on Women: A Systematic Review." *BMC Public Health*, 19(1): 158.

Sasson, Isaac. 2021. "Age and COVID-19 Mortality: A Comparison of Gompertz Doubling Time Across Countries and Causes of Death." *Demographic Research*, 44: 379–396. Publisher: Max-Planck-Gesellschaft zur Foerderung der Wissenschaften.

Spears, Dean. 2020. "Exposure to Open Defecation Can Account for the Indian Enigma of Child Height." *Journal of Development Economics*, 146: 102277.

Sun, Shiyi, and Jie Han. 2021. "Open Defecation and Squat Toilets, an Overlooked Risk of Fecal Transmission of Covid-19 and Other Pathogens in Developing Communities." *Environmental Chemistry Letters*, 19(2): 787–795.

UNICEF. 2022. "Community Sanitary Complexes in Rural India." UNICEF.

Vivek. 2023. "covid19bharat.org is Shutting Operations." Medium, accessed April 28, 2025.

Vyas, Sangita, and Nathan Franz. 2025. "Be Careful What You Ask For: Social Desirability Bias in Self-Reported Data in India." Social Science Research Network SSRN Scholarly Paper 4997356.

Vyas, Sangita, Nikhil Srivastav, Divya Mary, Neeta Goel, Sujatha Srinivasan, Ajaykumar Tannirkulam, Radu Ban, Dean Spears, and Diane Coffey. 2019.

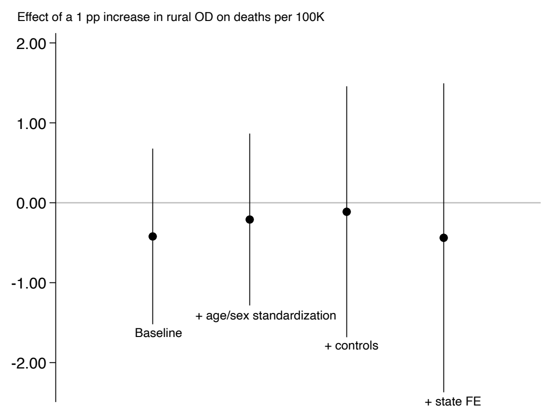
"Measuring Open Defecation in India using Survey Questions: Evidence from a Randomised Survey Experiment." *BMJ open*, 9(9): e030152.

World Health Organization. 2023. "WHO Coronavirus (COVID-19) Dashboard."

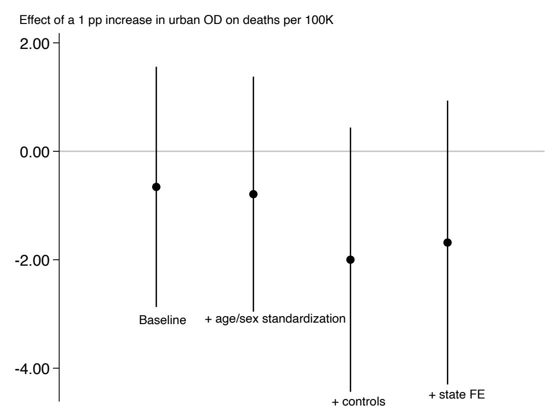
Supplemental Online Appendix, Not Intended for Publication

Appendix Figure A1

Open Defecation Rate and Annualized Excess Mortality Rate (Including Uttar Pradesh)



Panel A: Rural



Panel B: Urban

Notes: OD refers to open defecation. Whiskers indicate 95% confidence interval. Controls used are the household level controls indicate whether they are (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020.

Appendix Figure A2

Shared Toilets in India

Note: Source Kotwal et al. (2020). The image shows an example of a shared toilet in India, through shared toilet facilities can be widely different structurally and in how well they are maintained.

Appendix Table A1

Data Sources for Measures Used in the Paper

	Data availability			*State characteristic*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
State	All-cause mortality (vital records)	COVID-19 deaths (crowdsourced)	Individual mortality (DHS surveys from 2021)	Island?	City?	Union Territory?	Northeast?	Mountainous?
Andaman & Nicobar Islands		Yes	Yes	Yes		Yes		
Andhra Pradesh	Yes	Yes						
Arunachal Pradesh		Yes	Yes					
Assam		Yes						
Bihar	Yes	Yes						
Chandigarh		Yes	Yes	Yes				
Chhattisgarh		Yes	Yes					
Dadra & Nagar Haveli & Daman & Diu		Yes*				Yes		
Delhi		Yes	Yes	Yes				
Goa		Yes						
Gujarat		Yes						
Haryana	Yes	Yes	Yes					
Himachal Pradesh	Yes	Yes						Yes
Jammu & Kashmir		Yes	Yes					Yes
Jharkhand		Yes	Yes					
Karnataka	Yes	Yes						
Kerala		Yes						
Ladakh								
Lakshadweep		Yes	Yes	Yes		Yes		
Madhya Pradesh	Yes	Yes	Yes					
Maharashtra	Yes	Yes						
Manipur		Yes					Yes	
Meghalaya		Yes					Yes	
Mizoram		Yes					Yes	
Nagaland		Yes					Yes	
Odisha		Yes	Yes					
Puducherry		Yes	Yes	Yes		Yes		
Punjab		Yes	Yes					
Rajasthan	Yes	Yes	Yes					
Sikkim		Yes					Yes	
Tamil Nadu	Yes	Yes	Yes					
Telangana								
Tripura		Yes					Yes	
Uttar Pradesh	Yes	Yes	Yes					
Uttarakhand		Yes	Yes					Yes
West Bengal	Yes	Yes						

Notes: *Dadra & Nagar Haveli and Daman & Diu are included as one geographical entity in our analysis. All-cause mortality figures are used to construct excess mortality figures.

Appendix Table A2

State level Excess Mortality in India in Our Data and in Leffler et al. (2022)

(Different Time Periods)

	Excess Deaths per 100K[A]		
		Age-sex standardized	
			Leffler et al.[B]
State	(3)	(4)	(5)
Uttar Pradesh	36	36	41
Rajasthan	49	49	80
Himachal Pradesh	55	42	80
West Bengal	162	156	99
Bihar	158	168	203
Maharashtra	151	130	205
Karnataka	192	193	228
Haryana	187	175	291
Tamil Nadu	207	175	404
Madhya Pradesh	228	225	424
Andhra Pradesh	358	330	447
Assam	105	115	
Total	133	129	

Notes: [A]Excess deaths figures are annualized and calculated from mortality figures for the period of March 2020 to June 2021. [B]Excess deaths from Leffler et al. (2022) are based on January to June 2021 data and are not age-sex standardized. We construct and report annualized figures for excess deaths from their paper. Leffler et al. (2022) use 2020 data for Assam, so we omit their results here.

Appendix Table A3

State level Excess Mortality in India in Our Data and in Leffler et al. (2022)
(Same Time Period)

	Excess Deaths per 100K[A]		
	Age-sex standardized		Leffler et al.[B]
State	(3)	(4)	(5)
Uttar Pradesh	95	96	41
Rajasthan	118	117	80
Himachal Pradesh	86	66	80
West Bengal	214	207	99
Bihar	289	304	203
Maharashtra	286	247	205
Karnataka	330	331	228
Haryana	422	395	291
Tamil Nadu	124	105	404
Madhya Pradesh	611	600	424
Andhra Pradesh	588	539	447
Total	287	277	

Notes: [A]Excess deaths figures are annualized and calculated from mortality figures for the period of January to June 2021. [B]Excess deaths from Leffler et al. (2022) are also based on January to June 2021 data and are not age-sex standardized. We construct and report annualized figures for excess deaths from their paper. Assam does not have data for 2021, so is omitted in both ours and Leffler's analysis.

Appendix Table A4

Open Defecation Rate and Annualized All-cause Mortality Rate

	All-Cause Deaths per 100K					
				Age-sex standardized		
	(1)	(2)	(3)	(4)	(5)	(6)
Open Defecation (% households)						
Rural	-3.912***	-2.349	-0.573	-2.504**	-1.867	-0.097
	(1.002)	(1.456)	(1.338)	(0.987)	(1.422)	(1.304)
Urban	-0.640	-5.377**	-3.053	-1.980	-4.637**	-2.765
	(2.129)	(2.191)	(1.861)	(2.086)	(2.071)	(1.765)
Household covariates	No	Yes	Yes	No	Yes	Yes
State FE	No	No	Yes	No	No	Yes
Observations	241	241	240	241	241	240
Adjusted R-squared	0.067	0.368	0.613	0.041	0.333	0.593
Mean of outcome variable	625	625	625	602	602	602
SD of outcome variable	295	295	295	273	273	273
Mean of Rural open defecation	24.30	24.30	24.30	24.30	24.30	24.30
SD of Rural open defecation	17.45	17.45	17.45	17.45	17.45	17.45
Mean of Urban open defecation	6.95	6.95	6.95	6.95	6.95	6.95
SD of Urban open defecation	7.78	7.78	7.78	7.78	7.78	7.78

Notes: Controls are the district-level fraction (separately for rural and urban areas) of households that: (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020. Uttar Pradesh districts have been dropped from the analysis. Heteroskedasticity-robust standard errors, clustered by district, in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Table A5

Open Defecation Rate and Annualized COVID-19 Deaths

(Only Districts with Excess Mortality Data)

	COVID-19 Deaths per 100k people					
				Age-sex standardized		
	(1)	(2)	(3)	(4)	(5)	(6)
Open Defecation (% households)						
Rural	-0.472***	-0.363***	-0.234***	-0.380***	-0.302***	-0.183***
	(0.083)	(0.077)	(0.047)	(0.068)	(0.066)	(0.039)
Urban	0.341**	0.397***	0.155**	0.250**	0.319***	0.137**
	(0.152)	(0.132)	(0.075)	(0.122)	(0.114)	(0.063)
Household covariates	No	Yes	Yes	No	Yes	Yes
State FE	No	No	Yes	No	No	Yes
Observations	287	287	286	287	287	286
Adjusted R-squared	0.123	0.475	0.773	0.121	0.453	0.759
Mean of outcome variable	14.53	14.53	14.53	12.87	12.87	12.87
SD of outcome variable	24.88	24.88	24.88	20.45	20.45	20.45
Mean of Rural open defecation	26.56	26.56	26.56	26.56	26.56	26.56
SD of Rural open defecation	16.19	16.19	16.19	16.19	16.19	16.19
Mean of Urban open defecation	7.704	7.704	7.704	7.704	7.704	7.704
SD of Urban open defecation	7.961	7.961	7.961	7.961	7.961	7.961

Notes: Controls are the district-level fraction (separately for rural and urban areas) of households that: (i) are male-headed; (ii) are Muslim; (iii) identify as Scheduled caste (SC), Scheduled tribes (ST), Other Backward Classes (OBC), and no caste; (iv) have a household head with any education; (v) have a Below Poverty Line (BPL) card; and the district-level means of (vi) household size; (vii) number of rooms for sleeping; (viii) population density in 2020. Uttar Pradesh districts have been dropped from the analysis, since they are not in our main excess mortality sample. Heteroskedasticity-robust standard errors, clustered by district, in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.