Rainfall and death in a developing megacity*

Tom Bearpark† Archana Patankar ‡ Ashwin Rode§

July 26, 2024

A growing proportion of the world's population lives in developing cities, where rainfall routinely causes disruptive flooding. Linking all person-level, geolocated deaths in Mumbai, India, with sub-daily local rainfall measurements, we reveal that rainfall shocks cause a substantial portion of Mumbai's overall mortality, comparable to the death rate from all cancers. Impacts are concentrated amongst children, women, and slum residents, and are worse when heavy rainfall coincides with a high tide. These estimates reshape our understanding of the drivers of urban mortality, and have implications for urban planning policy and the projected costs of sea level rise and climate change.

*We thank the Princeton Center for Policy Research on Energy and the Environment and the University of Chicago Development Economics Research Fund for financial support. We thank Amin Gulamani, Setu Loomba, Harsh Vardhan Pachisia, and Smrithi Sharma for excellent research assistance. We are very grateful to Mook Bangalore, Ian Bolliger, Subrata Kumar Das, Eyal Frank, Michael Gechter, Subimal Ghosh, Simon Greenhill, Michael Greenstone, Dylan Hogan, Allan Hsiao, Maulik Jagnani, Amir Jina, Bob Kopp, Sulakshana Mahajan, Yogesh Mishra, Michael Oppenheimer, Manabendra Saharia, Tom Scott, Yixin Sun, Vaidehi Tandel, Shrabani Tripathy, Rachel Young, Guglielmo Zappalà, and workshop participants at the University of Chicago, Princeton University, Association of Environmental and Resource Economists, and Midwest International Economic Development Conference for their comments, suggestions, and help with data sources. Automatic Weather Station data was provided to us by the Brihanmumbai Municipal Corporation (BMC), tide tables were provided by the US National Oceanic and Atmospheric Administration (NOAA), and the Praja Foundation provided us with the disease data used in Appendix A.

†Princeton University, School of Public and International Affairs. e-mail: bearpark@princeton.edu

‡Green Globe Consulting. e-mail: archana.patankar09@gmail.com

§University of Chicago, Energy Policy Institute. e-mail: aarode@uchicago.edu

1 Introduction

Over 4.5 billion people live in cities, representing over half the world's population. The United Nations forecasts this number to grow to 6.7 billion by 2050, with nearly 90% of the increase taking place in Asia and Africa (UN DESA, 2018). This growing urban population is increasingly exposed to dangerous weather, a trend that will accelerate under future climate change (Tuholske et al., 2021; Rentschler et al., 2023). The combination of urbanisation and extreme weather poses significant socioeconomic challenges,[1] which may lead to increases in excess mortality (Picarelli et al., 2017; Guan et al., 2021). Yet these challenges remain poorly understood and understudied, especially with regard to cities in the developing world (Bryan et al., 2020). On one hand, greater wealth and access to infrastructure may protect urban populations from the harms of extreme weather. On the other, urban populations face unique public health burdens linked to high densities and sanitation, and also wide inequalities in who bears these burdens (Marx et al., 2013; Collier and Venables, 2017; Henderson and Turner, 2020; Kuddus et al., 2020).

Rainfall shocks, in which a large amount of rain falls in a short duration, are the dominant driver of urban flooding in many low- and middle-income cities. However, our understanding of the effects of rainfall and urban flooding is limited by the paucity of sufficiently granular data on both weather and human well-being (Jha et al., 2012; Gandhi et al., 2022). While a considerable body of work has focused on major flooding events including those caused by tropical cyclones (Elliott et al., 2015; Hsiang and Jina, 2014; Kocornik-Mina et al., 2020), the accumulated impact of smaller-scale, frequently occurring, and localised events is suspected to be comparably large but is rarely accounted for (UNISDR, 2013). In recent years, academic economics researchers (Guiteras et al., 2015; Donaldson and Storeygard, 2016; Chen et al., 2017; Patel, 2024) and disaster relief agencies (Schumann et al., 2023) have turned to satellite-based data sources to understand flooding and its impacts. However, it is widely recognised that densely built urban areas limit successful flood detection by satellites (Islam and Meng, 2022; Schumann et al., 2023).[2] Further, existing studies that focus on highly aggregated weather and outcome measures are neither able to represent specific hazardous weather conditions (e.g., the occurrence of heavy rainfall during a high

[1] These include inequality, crime, traffic congestion, communicable diseases, and environmental hazards (Henderson and Turner, 2020; Kuddus et al., 2020).

[2] There are several reasons for this. Satellite sensors have difficulty distinguishing between water and impervious smooth surfaces that are common in urban areas. The urban built environment also introduces confounding geometric factors such as layover and shadows. Additionally, features such as narrow streets limit the ground surface visible to the sensors. Hydrologists also face difficulties with modelling urban flooding due to its granular and often discontinuous spatial extent, shallow depth, and short-lived duration (Tanim et al., 2022).

tide period within a day) nor capture unequal harms across segments of an urban population. Thus, current research likely presents policy makers with a distorted view of the health impacts of urbanisation and extreme weather.

In this paper, we overcome these measurement challenges, and provide the first estimates of the mortality impacts of rainfall in an urban setting— Mumbai, India. We create a new dataset that links sub-daily rainfall measurements from a dense network of rain gauges with the universe of person-level mortality records identified by residential address. To causally identify the impacts of rainfall on mortality, we exploit random spatial variation in the incidence of rainfall shocks across the city within a given day— an artefact of Mumbai's climatology. Importantly, we also highlight the vastly unequal effect of these rainfall shocks based on person-level characteristics such as age, gender, and slum residence.

We find a steeply increasing relationship between rainfall and mortality. For instance, relative to a day with no rainfall, a day with 50mm of rainfall causes a 0.7% increase in mortality over the subsequent 5 weeks, while a day with 150 mm of rainfall causes a 2.2% increase in mortality over the same period.[3] The temporal profile of excess mortality from a high rainfall day exhibits strong effects both in the contemporaneous week and in subsequent weeks, suggesting that deaths may be caused both by immediate factors as well as by vector- and water-borne diseases that develop and spread over longer time horizons (Patankar, 2015). Our results imply that on average 7% to 10% of all monsoon season (i.e., June-September) deaths in Mumbai from 2006 to 2015 were caused by rainfall shocks— a share that is comparable to the city's overall share of deaths from all cancers. Given their magnitude, they reshape our understanding of the key drivers of urban mortality.

The harms from rainfall disproportionately affect vulnerable or disadvantaged groups, even within populations living in the same locality. Children under 5 years are the worst affected age group, with rainfall shocks explaining around 20% of their monsoon season deaths from 2006 to 2015. We find that women are more affected than men, consistent with literature that highlights gender inequality in the impacts of weather and climate change for both physiological and social reasons (Fruttero et al., 2023). Furthermore, starkly unequal effects are seen across slum and non-slum populations, which account for roughly 55% and 45% of the city's population respectively. A 150mm rainfall day causes a 2.9% increase in overall 5-week mortality for those residing in slums, but only a 1.1% increase for those not residing in slums.

[3]Between 2006 and 2015, there were an average of 27 days per year in which at least one location in Mumbai experienced 50mm or more of rainfall, and 5 days per year in which at least one location had more than 150mm (see Table 3 for more details).

Because urban flooding in Mumbai is mainly pluvial, an increase in rainfall can be expected to translate directly into increased flood hazard, all else being equal (Ranger et al., 2011). Although we are not able to directly observe flooding, two aspects of our results suggest that flooding is the key channel through which rainfall leads to excess mortality. First, we find that for both slum and non-slum populations, the mortality effects of extreme rainfall are greater in low-elevation localities than high-elevation localities. Second, we find that a high tide, which is known to impair the carrying capacity of the drainage system (Zope et al., 2015), greatly amplifies the mortality effects of rainfall, particularly when its timing within a day coincides with intense bursts of rain. The tide results suggest that in the absence of adaptive measures (e.g., drainage improvements), future sea level rise may substantially increase the mortality burden from rainfall shocks.

Mumbai offers a uniquely rich set of data sources and meteorological conditions that enable us to study the mortality impacts of rainfall shocks. To measure mortality, we utilise individual death records collected by Mumbai's governing civic body, the Brihanmumbai Municipal Corporation (BMC).[4] Importantly for this study, the records contain the date of death, age at death, gender, and residential address of each individual who died in Mumbai between 2006 and 2015. We use the address information in two ways. First we match each death with localised measures of rainfall from the BMC's network of automatic rain gauges located throughout the city, which measure rainfall in up to 15-minute intervals. Second, we classify each deceased person as either a slum resident or non-slum resident by matching the text in each residential address with census lists of slum names as well other keywords.

Our research design takes advantage of the unique meteorology of Mumbai's monsoon season, which lasts from June to September each year and accounts for almost all of the city's annual total rainfall of over 2,000 mm.[5] During this season, rainfall on a given day tends to occur in highly localised cloudbursts that drop heavy rain over some parts of the city while leaving other parts dry. Spatio-temporal variation in the city's daily rainfall has been described as "arbitrary" and "pattern-less", eluding prediction even by advanced statistical techniques (Singh et al., 2017).[6] This provides an opportunity to leverage quasi-

[4]The BMC's vital statistics registries are widely acknowledged to be some of the most advanced in India, and independent audits have validated the death records to be 95% complete (Mittra et al., 2021).

[5]For context, New York City gets around 1100mm of rainfall per year, which is spread roughly evenly across months of the year.

[6]Mumbai's susceptibility to extreme rainfall was most intensely demonstrated on July 26, 2005, when a record-setting 944 mm of rain fell over a 24-hour period at the airport weather station. However, even this remained localised, with rainfall totals over the same period varying widely across localities, from 810 mm at Bhandup and 490 mm at Dharavi in the northern and central parts of the city, to only 70 mm at Malabar Hill and Colaba in the southern part of the city. The interaction of synoptic-scale weather systems with coastal

random variation in rainfall across localities on a given day to identify a plausibly causal effect. In addition, we leverage exogenous variation in the timing of rainfall with respect to subdaily tidal fluctuations to isolate periods that pose the greatest risk of urban flooding.

Our results from Mumbai suggest that the global mortality burden of urban flooding is likely much higher than is currently understood. For example, the Emergency Events Database (EM-DAT), a widely used measure of disaster damages (Kahn, 2005; Cavallo et al., 2013; Gandhi et al., 2022) records fewer deaths due to floods for the whole of India, than we find for Mumbai alone during our sample period.[7] Like Mumbai, other developing megacities such as Manila, Jakarta, and Lagos face chronic problems with urban flooding, and thus likely also face a large and growing undocumented mortality burden from extreme rainfall. Further, climate change is likely to intensify monsoon season rainfall (Wang et al., 2021; Salunke et al., 2023), and will increase sea levels, both of which are key drivers of urban flooding. It is thus critical that urban planners and policy makers have high quality estimates of the magnitude of damages from urban flooding, and their distribution across vulnerable populations.

While this paper's primary contribution is to provide the first large-scale estimates of the unequal health impacts of recurrent rainfall shocks and urban flooding, it also builds on several strands of literature. Our work contributes to the broader literature on the challenges of urbanisation in developing countries, where weak and poorly enforced planning and zoning laws have meant that infrastructure often fails to keep pace with the growth in built environment and population. In particular, the differential mortality impacts of rainfall shocks for slum and non-slum residents highlight a possible channel through which improvements in housing (Cattaneo et al., 2009; Galiani et al., 2017), water, and sanitation (Cutler and Miller, 2005; Gamper-Rabindran et al., 2010; Alsan and Goldin, 2019; Ashraf et al., 2021; Bhalotra et al., 2021) can lead to better health outcomes.

Moreover, our work builds on the literature that studies the economic costs of environmental hazards. We contribute to the emerging literature on environmental inequality (Banzhaf et al., 2019; Cain et al., 2024) that uses population-wide data to characterise disparities in the causal effect of environmental hazards on health by person-level characteristics (e.g., by race in the US, as in Gillingham and Huang (2021) and Chakma et al.

land-surface features is thought to drive these rainfall idiosyncrasies across space (Singh et al., 2017).

[7] From 2006 to 2015, EM-DAT reports roughly 17,000 deaths in India due to 127 flood and storm events, most of which extend over large geographic areas (i.e., one or more states) (Centre for Research on the Epidemiology of Disasters, 2023). Only 11 of these events were identified to affect Maharashtra, the state in which Mumbai is located. In contrast, we find that local rainfall shocks caused over 22,000 deaths from 2006 to 2015 in Mumbai alone (Section 5).

(2024)).[8] We are the first to apply this approach to a developing country context when estimating disparities by slum vs. non-slum residence. Additionally, we extend the empirical literature on the impacts of weather on socioeconomic outcomes to the context of urban flooding. While this literature has extensively explored the mortality impacts of extreme temperature in various contexts (Burgess et al., 2017; Cohen and Dechezleprêtre, 2022; Carleton et al., 2022; Helo Sarmiento, 2023), the health impacts of extreme rainfall in urban settings have received almost no attention.[9] Besides quantifying the impacts, our results highlight potential mechanisms of adaptation (e.g., improved housing), which is a key focus of the research frontier on weather, climate change, and socioeconomic outcomes (Kolstad and Moore, 2020). We also develop a new approach to estimate how tidal fluctuations amplify the health impacts of rainfall shocks, thereby highlighting a new channel through which future sea level rise may affect human well-being. This complements a newly developing literature that uses spatial equilibrium models to quantify the economic costs of sea level rise (Desmet et al., 2021; Hsiao, 2023).

The rest of this paper is organised as follows. In Section 2 we describe our data, which includes the universe of deaths in Mumbai between 2006 and 2015 at the person-level, monsoon season rainfall from a dense network of urban weather stations at 15-minute intervals, and various covariates. In Section 3, we describe our approach to estimating the causal impacts of rainfall shocks on mortality, and uncovering heterogeneous effects by age, gender, slum residence, elevation, and tide patterns. In Section 4, we present our main empirical results on the effects of rainfall on mortality. In Section 5, we calculate the proportion of overall mortality in Mumbai that is caused by rainfall shocks. Finally, Section 6 concludes.

2 Data

To investigate the mortality costs of extreme rainfall and urban flooding, we compile a new dataset that links high resolution mortality and weather information. In conjunction with the mortality and weather data, we also make use of data on slum names and locations in

[8] This stands in contrast to other work that explores such disparities only by location-level, rather than person-level characteristics (e.g. Carleton et al. (2022)).

[9] Studies on rainfall have tended to focus on aggregate economic outcomes (Elliott et al., 2015; Damania et al., 2020; Holtermann, 2020; Russ, 2020; Kotz et al., 2022), health impacts in rural, agricultural settings (Maccini and Yang, 2009), or migration responses to major urban flooding events (Kocornik-Mina et al., 2020; Gandhi et al., 2022). Picarelli et al. (2017) on rainfall shocks and cholera incidence in Dar es Salaam, Tanzania is a rare example focusing on urban health impacts.

Mumbai, satellite-derived elevation and population data, and tide data.

2.1 Mortality data

Our mortality data are obtained through systematic queries of the Mumbai municipal government's deaths registration system website.[10] The resulting data provide us with the universe of recorded deaths in Mumbai between 2001 and 2015.[11] In our analysis, we use the deaths from 2006 to 2015 because granular weather data (Section 2.3) are available over those years.

For each death, the data provide a rich set of individual-level covariates, including the deceased's name, age, gender, residential address, and religion, but not cause of death. Our data are broadly consistent with Mumbai's city-level aggregate mortality statistics. For example, the total number of deaths in our data in 2015 is 91,797, while the BMC reports 94,706 in its official aggregate statistics.[12] Thus, in 2015, our deaths appear to capture 97% of the total deaths in Mumbai.

Using the information contained in the residential address, we locate each death to one of Mumbai's 89 PIN codes, which are India's postal area codes, analogous to US ZIP codes.[13] Figure 1 illustrates the stark spatial and temporal patterns in the data. Panel A shows the mean age of death in our mortality data between 2006 and 2015, mapped across Mumbai's PIN codes. There is extreme inequality across space, with some predominantly slum areas, such as Dharavi and the Mankhurd-Govandi belt, having a mean age of death of around 40.[14] In contrast, the mean age of death in wealthier areas is several decades higher; for instance, it is 73 in the affluent Breach Candy neighbourhood on the western seashore. The city-wide mean age of death in our 2006 to 2015 sample is around 57 years.

Panels B and C of Figure 1 illustrate the seasonality of mortality over months of the year.

[10] `https://portal.mcgm.gov.in/irj/portal/anonymous/qldldregreport?guest_user=english`, accessed in June 2024.

[11] Deaths from pre-2001 and post-2015 years are also available online, but the data quality is inconsistent during these periods.

[12] `https://www.mcgm.gov.in/irj/portal/anonymous/qlvitalstatsreport?guest_user=english`, accessed in July 2024.

[13] In our main analysis, we use the PIN code as the location identifier for a death. While this has the advantage of granularity, approximately 20% of the deaths do not report a PIN code that lies within Mumbai. This is either because 1) the deceased person's residential address was outside Mumbai (approx 5%), or 2) the deceased person's residential address was within Mumbai but was missing a PIN code (approx 15%). However, even in these cases, the data report in which of Mumbai's 24 municipal wards the death was registered. Further details on geo-location are in Appendix B.1. In Appendix D.8, we conduct robustness checks where we use the coarser, ward-level location identifier and retain observations lacking a PIN code.

[14] Dharavi is widely regarded as one of the world's largest slums (World Economic Forum, 2020).

During the monsoon months, the mean age of death in our data falls by around two years (Panel B). Panel C shows the number of deaths per month in our data broken out by age grouping— under 5 years, 5-64 years, and over 65 years. Seasonality is strongest in the youngest age group, for which there are more than 20% more deaths at the peak of the monsoon season compared to the rest of the year. This clear seasonality suggests to us that there is an important relationship between within-year climate and mortality.[15] However, given the potential for a wide range of confounding factors, we will need to leverage quasi-random variation in rainfall shocks to show there is a causal relationship underpinning this correlation.

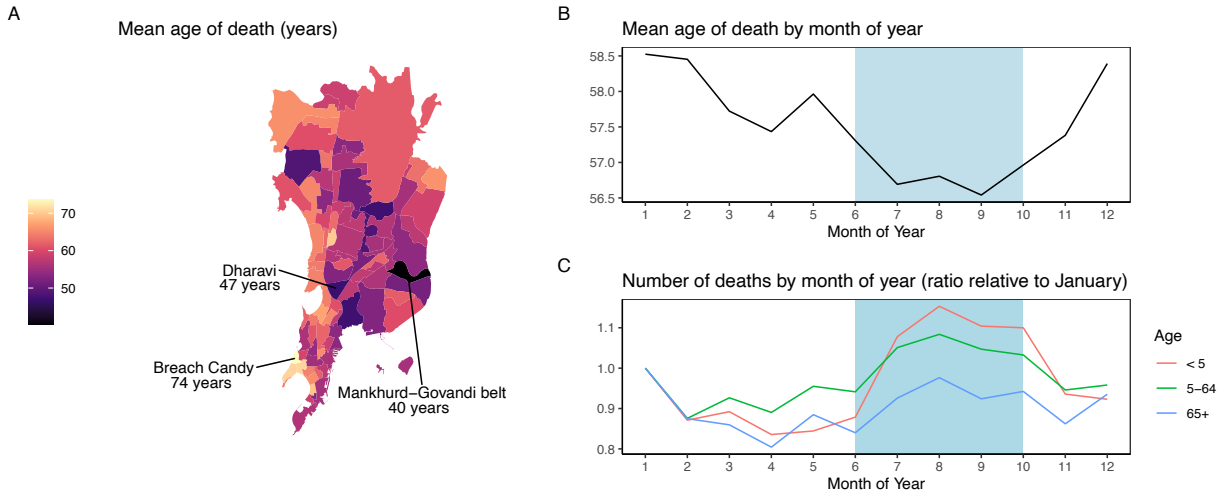

Figure 1: Seasonal and spatial patterns in mortality in Mumbai.
Panel A shows the spatial distribution of the mean age of death between 2006 and 2015 across PIN codes in Mumbai. We highlight the PIN codes containing two of Mumbai's biggest slum areas, Dharavi (400017) and the Mankhurd-Govandi belt (400043), as well as the affluent Breach Candy neighbourhood (400026). Panel B shows the mean age of death across months of the year in our sample, with the monsoon season shaded in blue. We can see clear visual evidence of seasonality. Panel C shows the number deaths by months of the year, broken out by age category, and normalised to be equal to one in January for each age category.

The deaths data also reveal variation and inequities along other dimensions.[16] However, the city-wide mean age of death steadily increased over the sample period, from 53 in 2006 to 61 in 2015, reflecting general improvements in health care and hygiene.[17] Furthermore,

[15]The seasonality in our deaths data is also mirrored in data on the incidence of diseases, especially those such as dengue, diarrhoea, malaria and typhoid which are known to be linked to urban flooding (Patankar, 2015). See Appendix A.

[16]For example, the mean age of death for women is around 5 years higher than that for men. The mean age of death for Hindus is nearly ten years higher than that of Muslims.

[17]This increase is consistent with the trend in life expectancy for India as a whole, which increased from 65 years in 2006 to 70 years in 2015. (See `https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=IN`.)

an increasing trend in the mean age of death is found in 87 out of the 89 PIN codes.[18]

2.2 Slum classification

Our mortality data contains detailed person-level residential address information. We use this information to classify whether the deceased likely lived in a slum, allowing us to create separate time series of death counts for slum and non-slum populations in each PIN code. To implement the classification, we use a combination of fuzzy string matching and text mining approaches. Specifically, we obtain a list of names of the slums in each of Mumbai's 24 municipal wards from the 2001 census, and match these into the addresses that fall in that ward using a fuzzy matching procedure. We supplement this matching procedure using a search for terms that are clearly associated with slum addresses.[19] Overall, our procedure categorises around 55% of our addresses as slum addresses, which is consistent with aggregate statistics from the 2001 census as well as other more recent estimates (WRI-India, 2022). Further details of this process, and our validation procedures are in Appendix B.2.

Panel A of Figure 2 shows the spatial pattern of the percent of deaths that we categorise as slums by PIN code. These patterns are closely related to the patterns in the mean age of death shown in Figure 1A. In Panel B of Figure 2, we scatter the mean age of death against the percentage of slum deaths. As expected, we find a negative relationship. PIN codes which have a greater proportion of deaths in slums also have a lower mean age of death.

[18]The two PIN codes in which the mean age of death is not increasing have statistically insignificant trends.

[19]Specifically, the terms are *zopadpatti* (the Marathi language word for slum, including abbreviated forms *zop*, *zp*, and *z.p*), *chawl* (a type of dense of communal housing, including abbreviated form *chl*), "police" (addresses of homeless persons or residents of informal settlements may be recorded as the nearest police station), *nagar* (literally a district or quarter within a town, but in Mumbai commonly applied to slum settlements), and "transit". The word "transit" is included to capture "transit camps", which are temporary settlements for persons displaced due to construction projects or from dilapidated buildings about to be destroyed or redeveloped. In practice, however, transit camps can persist as long-term settlements.

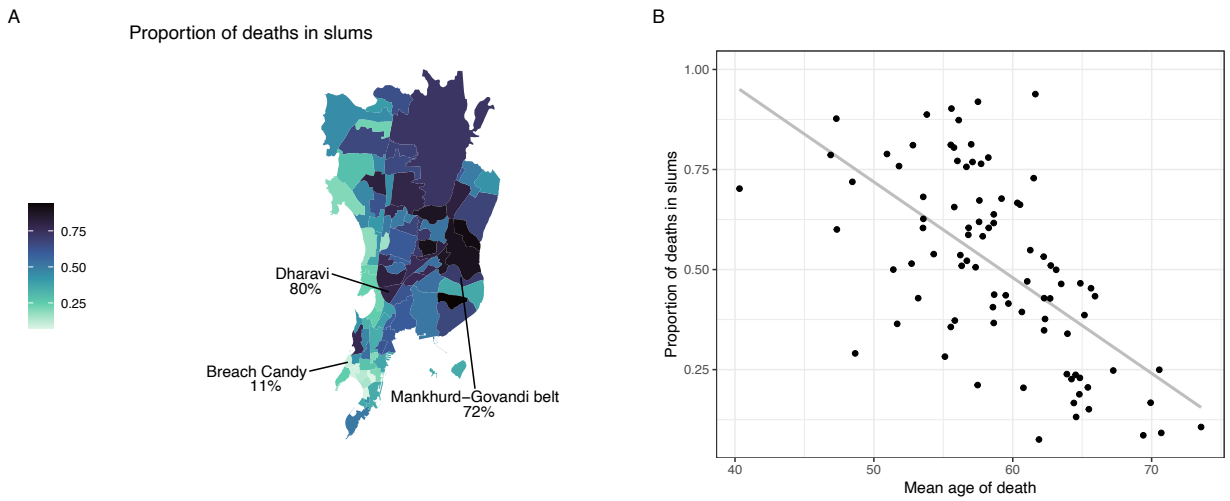

Figure 2: Proportion of deaths in slums and mean age of death, by PIN code.
Maps shows the spatial distribution of the percent of 2006-2015 deaths in slums across PIN codes in Mumbai. We highlight the PIN codes containing two of Mumbai's biggest slum areas, Dharavi (400017) and the Mankhurd-Govandi belt (400043), as well as the affluent Breach Candy neighbourhood (400026). The scatter plot displays the association between the percent of 2006-2015 deaths in slums and the mean age of death for 2006-2015 deaths, with each point representing a PIN code.

2.3 Weather data

Our main weather data are rainfall measurements collected from a dense network of Automatic Weather Stations (AWS), which were set up by the BMC after the July 2005 floods. These stations provide rainfall totals at 15-minute intervals, for an unbalanced panel of stations across Mumbai. The number of stations range from 22 in 2006 to 51 in 2015.[20] The AWS data also include temperature measurements, which we use as a control variable in a robustness check. To aggregate these 15 minute interval station level data to the day-by-PIN code level, we first create a balanced panel of station level data within each year, imputing missing values across space (Auffhammer et al., 2013). We then calculate PIN code level averages of the available stations in each year, weighting the stations by the inverse of the square of the distance between each PIN code's centroid and the location of the stations. Further details of the dataset cleaning and construction procedure, and the station availability, are in Appendix B.3.

A high level of spatial resolution is crucial for capturing the idiosyncratic patterns of rain-

[20]In most years, the AWS data are available only for the monsoon season (See Appendix Table B.3). Details of the AWS data availability are in Appendix B.4. In Appendix D.10 we show robustness of our main results to creating our climate data measures using only a balanced sample of stations that are available to us in all years.

fall in Mumbai. On a given day, rainfall in the city occurs in highly localised bursts, and this quasi-random variation allows us to credibly identify causal impacts on mortality. Table 1 illustrates the identifying variation used in our analysis by comparing the standard deviation of PIN code-level daily rainfall observations before and after conditioning out spatial and temporal fixed effects. As a baseline, $\hat{\sigma}_{it}$ denotes the raw standard deviation of the actual PIN code-level daily rainfall observations. We then condition out PIN code $\times$ year fixed effects from these observations, and calculate the standard deviation of the residual variation, which we refer to as the "within PIN code-year SD". Finally, we further condition out date fixed effects,[21], and refer to the residual standard deviation from this model, which mimics our estimation strategy, as "full residual SD". The final row presents the ratio of this residual standard deviation to the raw standard deviation in the daily rainfall values, $\hat{\sigma}_{it}$. Even after removing extremely saturated fixed effects that flexibly control for date-specific seasonality and PIN code-specific inter-annual trends, we still retain substantial variation that can be used for identification (i.e., roughly 30% of the original raw variation).

Quantity	Estimate
$\hat{\sigma}_{it}$	26.2
Within PIN code-Year SD	25.6
Full residual SD	7.5
Ratio of full residual SD to $\hat{\sigma}_{it}$	0.29

Table 1: Variation in daily rainfall.
Table shows summary statistics for PIN code-level daily rainfall observations in our sample. Each row presents a measure of the standard deviation of PIN code-level daily rainfall. The first row shows the raw standard deviation. The second presents residual standard deviation after conditioning out PIN code $\times$ year fixed effects, and the third the residual standard deviation after conditioning out PIN code $\times$ year fixed effects and date fixed effects. The last row shows the ratio of the residual standard deviation in the penultimate row to the raw standard deviation in the first row.

Figure 3 depicts the variation used in our study by mapping the spatial distribution of daily rainfall in Mumbai on individual days. Specifically, we map PIN code level daily rainfall values for the two days in our sample containing the highest rainfall values, as well as days immediately preceding and following these two days.

[21]For instance, July 1, 2014.

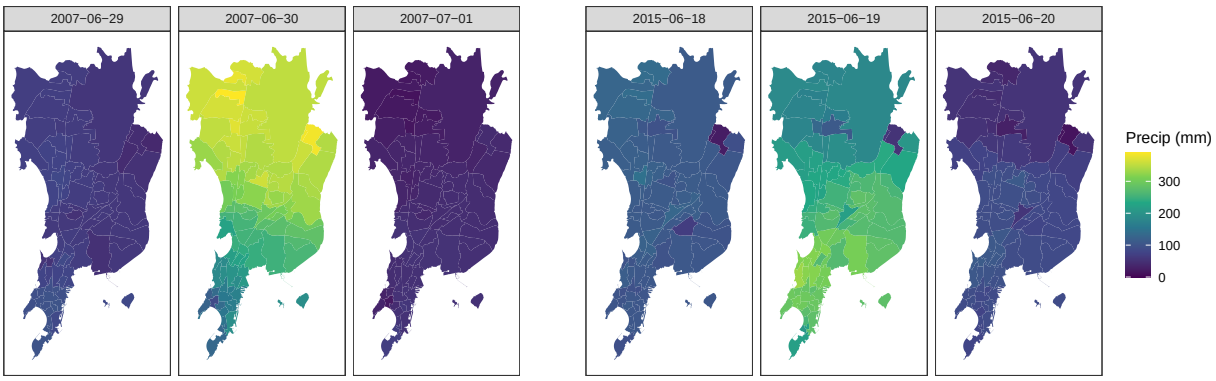

Figure 3: Total daily rainfall on two extreme days.
Maps illustrate the spatial distribution of total daily rainfall across PIN codes of Mumbai for the two days containing the highest PIN code × day rainfall observations in the sample (i.e., June 30, 2007, and June 19, 2015), and the days immediately preceding and following these days.

Figure 3 demonstrates that rainfall is highly temporally and spatially localised. Even on very extreme rainfall days, there are some areas of the city that receive little rainfall. This is consistent with other studies (Singh et al., 2017) which have found that spatial patterns of rainfall in Mumbai are extremely idiosyncratic and hard to forecast. Despite this high level of idiosyncrasy, the yearly total rainfall experienced by each PIN code is roughly uniform (see Appendix B.3.1).

2.4 Other data

We supplement our mortality and weather data with a range of other datasets that include variables such as elevation, population, and tides.

We collect elevation data at a resolution of roughly 50 m × 50 m, using the R package elevatr[22], which is an interface to the Mapzen terrain tiles. We use these data to calculate the average elevation for each of Mumbai's PIN codes, and the proportion of each PIN code that is below 5 m elevation. Low elevation PIN codes are expected to be at higher risk of urban flooding.[23]

[22] `https://github.com/jhollist/elevatr`

[23] In an additional robustness check (Appendix E.4), we also use the "Height Above Nearest Drainage" (HAND) rather than elevation to categorise PIN-codes. HAND is defined as the the vertical distance between a location and its nearest stream, and we calculate the average HAND for each PIN code using a

To calculate the hourly tide level for Mumbai, we use historical tide tables from the US National Oceanographic and Atmospheric Administration.[24] These provide us with the high and low tide predictions for each day in our sample.[25] We smoothly interpolate these values to obtain predictions at the hourly level and match these to our sub-daily rainfall data, in order to test whether rainfall is more damaging when it coincides with a high tide. Figure 4 presents time series of the interpolated tide values for August 2009. Alongside the level of the tide in each hour of our sample, we calculate two other statistics that are used in robustness checks in Appendix E.5. Specifically, we record whether the maximum tide in each hour of our sample is above or below 450cm, since BMC documents claim that flooding is especially likely when the tide is above 450cm.[26] We also make use of the average tide value within a day, which is shown in the right panel of Figure 4.

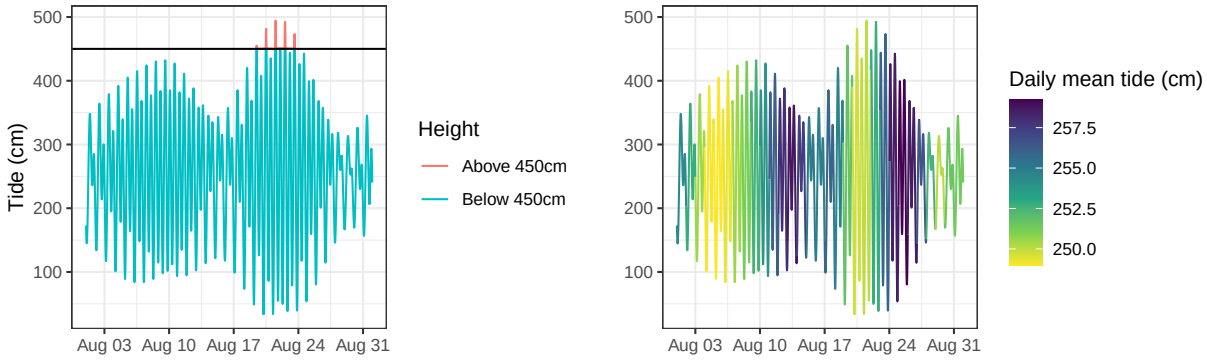

Figure 4: Descriptive tide plots.
Plots show time series of Mumbai's predicted tide patterns across August 2008. Data are interpolated from the highs and lows to the hourly level using a smooth spline interpolation. Panel (A) highlights the portion of the month during which the tide prediction is above 450cm. Panel (B) colours the time series by the average tide level across the 24 hours in each day.

We collect a variety of additional data for use in sensitivity and robustness checks. The Praja Foundation provided us with disease data, which allow us to characterise seasonal patterns in diseases (Appendix A). We obtained daily and subdaily data on temperature

HAND map for Mumbai developed in Tripathy et al. (2020), which we obtained through direct correspondence.

[24]Obtained through correspondence, but post-2008 tables are available here: `https://tidesandcurrents.noaa.gov/historic_tide_tables.html`

[25]We only have one time series of tide data available to us, which is from a single tide-gauge located at the southern tip of the city. However, independent measurements have confirmed that tide height and phase are similar at different points along Mumbai's coastline (National Institute of Oceanography, 2016).

[26]A broad range of BMC documents, NGO reports, and news alerts state a threshold tide level of 450 cm, over which acute flooding is very likely to occur. See for example `https://portal.mcgm.gov.in/irj/go/km/docs/documents/Coastal%20road/Coastal%20Road%20DPR/EIA%20report.pdf`.

and rainfall from the Indian Meteorological Department's (IMD) two weather stations in Mumbai, and use these to validate the AWS data (Appendix B.3). We obtained a shapefile used by Tandel et al. (2022) for the locations of slums in Mumbai through direct correspondence with the authors (Appendix C.2), which we use to characterise the elevation of slum areas within a PIN code. Finally, we construct PIN code-level population estimates, using the Global Human Settlement Layer (GHSL) satellite-derived data product (Schiavina et al., 2023), which provides population estimates every five years in our sample at a resolution of roughly 100 m $\times$ 100 m. This allows us to conduct robustness checks in which we population weight our regressions (Appendix D.6).

2.5 Regression data summary statistics

Table 2 presents summary statistics for the data used in our main analysis.

	Mean	SD	Median	Min	q^{99}	Max	Total Deaths
Total Deaths	15.8	11.5	13.0	0	51.0	128.0	363380
Deaths below 5 yrs	1.1	1.9	0.0	0	8.0	26.0	24853
Deaths 5-64 yrs	7.2	6.1	6.0	0	27.0	105.0	165710
Deaths 65 yrs and above	7.4	5.4	7.0	0	23.0	39.0	169674
Male deaths	9.0	7.0	8.0	0	31.0	107.0	207593
Female deaths	6.8	5.3	6.0	0	23.0	36.0	155744
Slum deaths	8.6	8.6	6.0	0	37.0	63.0	197577
Non-Slum deaths	7.2	5.9	6.0	0	25.0	93.0	165803
Daily rainfall total	11.4	26.2	0.7	0	130.0	389.9	
Weekly rainfall total	80.3	119.9	23.3	0	520.2	766.2	

Table 2: Summary of data used in analysis.
Table shows summary statistics for PIN code-level weekly death totals and weekly deaths by age group, gender, and slum residence. We also show PIN code level daily and weekly rainfall totals. The 99^{th} percentiles are indicated by q^{99}. Totals deaths summed across age groups and or genders are not exactly equal to Total Deaths in the first row, due to missing age or gender information in a small number of observations.

On average, PIN codes have about 16 deaths per week during our sample period. However, there is substantial variance across time and space. Roughly 1 out of every 16 deaths in our sample is for someone under the age of 5.

Similarly, our treatment variable, rainfall exhibits substantial variation over the estimating

sample. At the daily level, the mean PIN code level rainfall is about 11mm.[27] However, the distribution is highly skewed, with each year containing a substantial number of PIN code $\times$ days with extreme rainfall values (Table 3). We exploit this variation in rainfall across PIN codes at the daily level to estimate a nonlinear effect of daily rainfall on mortality.

Year	N	$\geq$ 50mm	$\geq$ 100mm	$\geq$ 150mm	$\geq$ 200mm	$\geq$ 300mm
2006	17177	1160	539	191	35	0
2007	32485	1166	389	168	67	40
2008	24386	1451	319	55	1	0
2009	24386	821	309	96	12	0
2010	13528	1732	406	39	3	0
2011	14952	1216	579	167	8	0
2012	8188	530	87	4	0	0
2013	5340	1088	420	95	12	0
2014	13528	1268	362	33	0	0
2015	10769	766	272	89	75	13

Table 3: Extreme rainfall observations by year.
Table shows the total number of PIN code $\times$ days (N) observed in each year in our AWS data, and the number of PIN code $\times$ days in each year with rainfall above 50 mm, 100 mm, 150mm, 200 mm, or 300 mm. Details on the sample available to us in each year are in Appendix B.4.

3 Estimating a mortality-rainfall response function

3.1 Empirical strategy using daily rainfall totals

Here we present the empirical strategy and regression equation that we use to estimate the mortality impact of rainfall in Mumbai. We exploit quasi-random variation in the location and timing of daily rainfall to estimate the following model on weekly mortality data for the 89 PIN codes within Mumbai:

$$\underbrace{m_{i,w,y}}_{\substack{\text{Log expected deaths in pincode } i \\ \text{Week of year } w\text{, year } y}} = \underbrace{f(\boldsymbol{R}_{i,w,y})}_{\substack{\text{Rainfall response} \\ \text{function}}} + \underbrace{\alpha_{i,y} + \lambda_{w,y}}_{\text{Fixed effects}}, \tag{1}$$

where i indexes the PIN code, w indexes the week of the year, and y indexes the calendar year. The outcome variable, which is the expected number of log deaths in a PIN code

[27] This is the average during our estimation sample, which is usually only the monsoon season (Section 2.3).

$\times$ week, is modelled as a linear-in-parameters function f of a vector of rainfall variables, $\boldsymbol{R}_{i,w,y}$, and a granular set of location and time fixed effects. Since the number of deaths in a PIN code $\times$ week is a non-negative count variable, we estimate the parameters in Equation (1) using Poisson pseudo maximum likelihood (Chen and Roth, 2024).

The mortality-rainfall response function f causally relates daily rainfall to mortality outcomes. There are two key features of f worth highlighting. First, it allows for non-linearity in the response of weekly mortality to daily rainfall. Specifically, we construct non-linear transformations (e.g., polynomials) of daily rainfall before summing these over days of the week, which allows the function f to capture the impacts of the distribution of extreme days within the week (Hsiang, 2016). Second, the arguments of f include lagged values of the vector of rainfall variables. This is important because rainfall in a given week may affect mortality in subsequent weeks. For instance, an increase in mortality due to extreme rainfall in a given week could be compensated for by a fall in mortality in subsequent weeks. This could occur if extreme rainfall causes a deterioration in the health of individuals who would have died even in the absence of the rainfall shock. In this case, the only effect of the rainfall shock is to alter the timing of mortality by a few weeks, but not the number of deaths in the longer run.[28] Alternatively, extreme rainfall in a given week may lead to increased mortality in subsequent weeks, because (for example) water-borne diseases take time to develop and spread. In this case, the rainfall shock has persistent effects that result in a higher number of deaths in the longer run. Including lagged values in f accommodates both these possible narratives.

In our main specification, we model the mortality-rainfall response function using a third-order polynomial in daily rainfall, for days in the contemporaneous week and 4 lagged weeks.[29] Letting $R^k_{i,d,w,y}$ represent the total daily rainfall on day d raised to the power k, the function f is constructed as follows:

$$f(\boldsymbol{R}_{i,w,y}) = \sum_{k=1}^{3} \sum_{l=0}^{4} \beta_{l,k} \sum_{d \in w-l} \left(R^k_{i,d,w-l,y} \right), \tag{2}$$

where each parameter, $\beta_{l,k}$, denotes a coefficient on the k^{th} power and l^{th} lag term.

[28] This type of temporal displacement is sometimes referred to as a "harvesting" effect (Deschênes and Moretti, 2009).

[29] We emphasize results from the third-order polynomial because it captures important nonlinearities while being relatively parsimonious. We emphasize results from a model with the contemporaneous week and 4 lagged weeks because it captures temporal displacements while allowing us to retain weeks in the estimating sample that would be lost if additional lagged were to be included. In Appendix , we show results under alternative functional forms (Appendix $D.1$) and number of lags (Appendix $D.4.1$).

Apart from the function f, the fixed effects are a crucial part of Equation (1). In our main specification, we use both PIN code $\times$ year fixed effects, $\alpha_{i,y}$, and week-by-year (i.e., date) fixed effects, $\lambda_{w,y}$. The PIN code $\times$ year fixed effects flexibly control for any annually-varying PIN code-specific variables,[30] while the date fixed effects flexibly control for any city-wide shocks in a given week as well as within-year seasonality. Together, these fixed effects imply that we are identifying the mortality-rainfall response function from idiosyncratic spatial variation in rainfall within the city in a given time period. We assume that conditional on the included fixed effects, rainfall is as good as randomly assigned, allowing for a causal interpretation of the function f.[31] This research design is possible due to the highly localised nature of where rain falls on a given day in Mumbai, and using high resolution rainfall data is thus crucial for the analysis. In Appendix D.9, we show results under alternative specifications of the fixed effects that are even more saturated, and find qualitatively similar results.

Throughout our analysis, we use a conservative two-way clustered standard error estimator, that allows for arbitrary correlation of residuals across weeks within a given PIN code, and across PIN codes within a week.

We also consider specifications that allow for heterogeneity in the mortality-rainfall response based on individual-level characteristics. To do so, we estimate a separate response for the set of individuals in each group a:

$$m_{a,i,w,y} = f_a(\boldsymbol{R}_{i,w,y}) + \alpha_{a,i,y} + \lambda_{a,w,y}. \tag{3}$$

Equation (3) is estimated using a stacked regression, with all elements of the rainfall vector $\boldsymbol{R}_{i,w,y}$ and all fixed effects interacted with group-level indicator variables. We consider groupings based on an individual's age, gender, and type of residence. For age, we classify individuals into three categories, $a \in \{< 5 \text{ years}, [5, 65) \text{ years}, \geq 65 \text{ years}\}$, and we classify each individual's residence as slum or non-slum based on residential address (See Section 2.2).

[30] One such variable is the set of weather stations available to us in each year. Because the AWS weather data is an unbalanced panel, the PIN code $\times$ year fixed effects are needed to control for the varying composition of stations in each year.

[31] Note, our estimator further assumes that there are no spillover effects. We discuss this assumption, and the possibility of both spatial and temporal spillovers in Appendix D.4. Overall, we find do not find evidence that un-modelled spillovers significantly bias our results.

3.2 Within-day timing of rainfall

Our main analysis focuses on the effects of daily rainfall shocks on mortality. In particular, the specifications in Equations (1) and (3) account for the nonlinear effects of daily rainfall on weekly mortality. While daily rainfall is a plausibly salient variable for public and private decision making, the native resolution of our AWS rainfall data is the 15 minute level. This allows us to utilise variation in the timing of rainfall within a day to investigate the effect of extreme and short-lived rainfall shocks on mortality. For this purpose, we consider a specification where we construct non-linear transformations of rainfall at the hourly level, before aggregating to the weekly level:

$$m_{i,w,y} = g(\boldsymbol{R}^{hourly}_{i,w,y}) + \alpha_{i,y} + \lambda_{w,y}. \tag{4}$$

Letting $R^k_{i,h,d,w,y}$ represent the total rainfall on hour h of day d raised to the power k, the function g is constructed as follows:

$$g(\boldsymbol{R}^{hourly}_{i,w,y}) = \sum_{k=1}^{3}\sum_{l=0}^{4} \beta^{hourly}_{l,k} \sum_{d\in w-l}\sum_{h\in d} \left(R^k_{i,h,d,w-l,y}\right), \tag{5}$$

where each parameter, $\beta^{hourly}_{l,k}$, denotes a coefficient on the k^{th} power and l^{th} weekly lag term. The function g here thus captures the nonlinear effects of hourly rainfall on weekly mortality.

Furthermore, we also exploit exogenous variation in tide height to estimate a specification that allows rainfall to have different impacts, depending on the level of the predicted maximum tide in the hour that it falls. Let $\tau_{h,d,w,y}$ denote the hourly tide level on hour h of day d in week w and year y. Specifically, we construct the function g_τ, which augments the function g by adding multiplicative interactions of each hourly rainfall polynomial term with hourly maximum tide height:

$$g_\tau(\boldsymbol{R}^{hourly}_{i,w,y}) = g(\boldsymbol{R}^{hourly}_{i,w,y}) + \sum_{k=1}^{3}\sum_{l=0}^{4} \gamma^{hourly}_{l,k} \sum_{d\in w-l}\sum_{h\in d} \left(R^k_{i,h,d,w-l,y}\cdot \tau_{h,d,w-l,y}\right), \tag{6}$$

In Appendix E.5, we consider two additional specifications that leverage variation in tide patterns. First, we consider a model in which hourly rainfall is interacted with an indicator for whether the tide is above or below 450cm. A broad range of BMC documents, NGO reports, and news alerts state a threshold tide level of 450 cm, over which acute flooding is

very likely to occur.[32] Second, we consider a model in which we interact the daily average tide level with the rainfall response function from Equation (2). Overall, these specifications build evidence that tide is a key determinant of the mortality-rainfall relationship, and thus that urban flooding is likely to be a important mechanism.

4 Results

4.1 The mortality-rainfall response function

Figure 5 presents the results from estimating the causal response function from Equation (1), using a third-order polynomial in daily rainfall for days in the contemporaneous week and 4 lagged weeks. In all figures, a day with zero total rainfall is treated as a reference. Therefore, points on the curve represent the percent impact of a day with a level of rainfall shown on the x-axis on the 5-week mortality rate, relative to if that day had zero rainfall.[33]

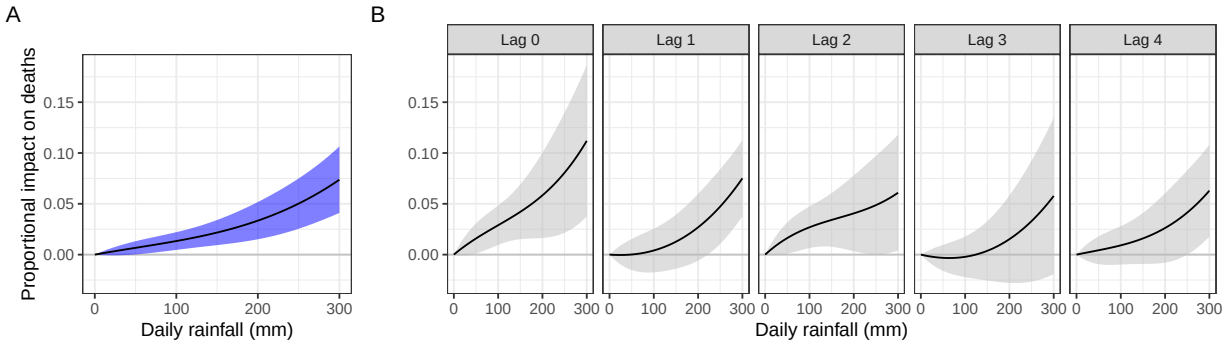

Figure 5: Percent change in 5-week mortality due to daily rainfall.
Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall for days in the contemporaneous week and 4 lagged weeks. Shaded areas indicate 95% confidence intervals. In Panel A, Points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. In Panel B, points along each curve represent the percent change in weekly mortality in the contemporaneous week (leftmost plot) and first to fourth subsequent weeks (second-from-left plot to rightmost plot, respectively), due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall.

[32] See for example `https://portal.mcgm.gov.in/irj/go/km/docs/documents/Coastal%20road/Coastal%20Road%20DPR/EIA%20report.pdf`.

[33] Specifically, the value of the curve at a given daily rainfall value R is obtained by multiplying each coefficient $\beta_{l,k}$ by R^k (Equation (2)), summing over these products, and then dividing by 5. The division by 5 is done to account for the fact that the deaths caused by a day with rainfall R occur over a 5-week period; the division converts a percent increase in weekly deaths to a percent increase in deaths over a 5-week period.

Panel A shows that on average across our sample, mortality increases sharply with daily rainfall.[34] For instance, a single day with 150 mm rainfall causes an increase in 5-week mortality by around 2.2%. This result stands in sharp contrast to estimates of the mortality impacts of temperature shocks in urban India, which found that hot days do not have an effect on urban residents (Burgess et al., 2017).

The response depicted in Panel A represents the effect of a single day with a given amount of rainfall on mortality over a 5-week period. In Panel B, we present the decomposition of the cumulative effect shown in Panel A, which gives evidence of persistent excess mortality over this entire period. Excess mortality from a rainy day is found to be roughly equal in the contemporaneous week and each of the 4 subsequent weeks. This is consistent with a narrative that beyond the immediate mortality repercussions of extreme rainfall in the contemporaneous week, flood-related diseases take time to develop and spread. We do not find evidence that rainfall shocks simply advance the timing of deaths without altering the total number of deaths.

4.1.1 Robustness, sensitivity, and falsification checks

We present an extensive suite of robustness checks in Appendix D. We show our results are robust to estimating at the daily level with 35 lags (Appendix D.1), various choices of functional forms for f (Appendix D.2) and to specifications where nonlinear transformations of rainfall are constructed at the hourly or weekly level, rather than at the daily level (Appendix D.3).

We also consider a range of specifications that investigate the role of temporal and spatial spillovers. In Appendix D.4.1, we estimate regressions of daily mortality on daily rainfall with up to 95 lags. These results suggest that including rainfall for only the past 35 days or 5 weeks likely leads to an underestimate, as mortality effects persist for a much longer duration. However, including more lags reduces the number of observations available to us and results in less statistical power, especially when estimating separate responses by individual characteristics. We therefore focus on a model with 5 weeks of rainfall, noting that it likely understates the full extent of the mortality impact. In Appendix D.4.2, we estimate specifications at different levels of spatial aggregation to account for the possibility that the effect of rainfall shocks is spread over a larger area than the PIN code (e.g., through flow of water, or movement of people). Overall, we do not find meaningful evidence that

[34] These and other results from this section are presented in tabular form in Appendix C.1. It is important to note that days with a much higher rainfall than 300 mm occur in Mumbai, and that our estimates are precise even at the extremes of our estimating sample.

spillover effects substantially bias our main results.

Additionally, we show that our estimated mortality-rainfall response function is qualitatively similar when varying the estimating sample (Appendix D.5), when estimating with Poisson, ordinary least squares, and population-weighted least squares regressions (Appendix D.6), when we include PIN code-level temperature controls (Appendix D.7),[35] and when using different strategies to assign deaths to locations (Appendix D.8).[36] We also show that our results are robust to using even more saturated fixed effects than in our main specification (Appendix D.9). Importantly, in Appendix D.10, we conduct robustness checks where we make alternative decisions on the cleaning steps involved in converting our raw AWS station level data to PIN code-by-day summary statistics. Specifically, we show our results are robust to using a cleaning procedure where we only use a subset of the AWS stations that are available to us in every year of our sample.

Finally, in Appendix D.11 we conduct a series of randomisation placebo checks to show that spurious correlations are not driving our results. We fail to find evidence of a strong rainfall-mortality relationship when randomly shuffling rainfall values either across dates and PIN codes, across dates within each PIN code, or across PIN codes within each date.

4.2 Heterogeneity in the mortality-rainfall response function across age group and gender

Next, we investigate heterogeneity in the mortality-rainfall response function according to age-group and gender. To do so, we create separate time series for each PIN-code for deaths in each age category and gender respectively, and then allow the effect of rainfall to vary according to Equation (3).

[35] Among the variables the date fixed effects in Equation (1) control for is city-level average temperature in a given week. Because temperature in Mumbai is much more spatially uniform than rainfall, these date fixed effects are likely not appropriate for analysing the causal impacts of temperature as they absorb a great deal of the relevant variation. Nevertheless, in Appendix D.7, we conduct a robustness check where we control for PIN code-specific daily temperatures within a week while also including the $\lambda_{w,y}$ fixed effects. We find that the mortality-rainfall response is robust to the inclusion of temperature controls, and that the mortality effects of temperature are imprecisely estimated.

[36] In our main specification, we assign each death to the PIN code listed in the residential address. This requires us to throw out roughly 20% our data which do not include PIN codes, or include PIN codes that are not within the city of Mumbai. However, we can alternatively assign deaths to municipal wards (a coarser location than PIN codes) using the ward identifier in our raw data.

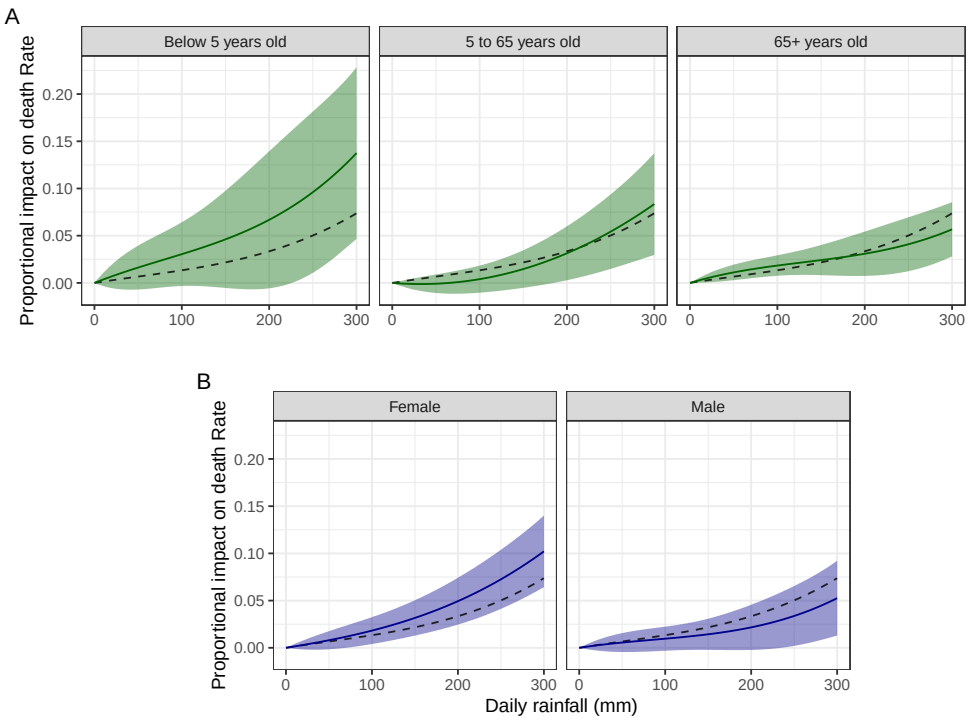

Figure 6: Percent change in 5-week mortality due to daily rainfall, heterogeneity by age group and gender.

Estimates are from models that uses a third-order polynomial in daily rainfall for days in the contemporaneous week and 4 lagged weeks. Points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. Panel A presents estimates from a version of Equation (3) allowing for heterogeneity across age categories. Panel B presents estimates from a version of Equation (3) allowing for heterogeneity across gender categories. In each plot, the average response function from Figure 5 Panel A is reproduced as a dashed black line for reference.

Figure 6 Panel A allows for heterogeneous impacts of rainfall on mortality by age category (< 5 years, 5-65 years, ≥ 65 years). We select our age categorisation based on previous climate impacts literature (Carleton et al., 2022), and because we find these capture the relevant heterogeneity well (Appendix E.1). We find that the youngest age group is the most vulnerable to the mortality impacts of rainfall, followed by the oldest age group.[37] The huge estimated impacts for the youngest age group are especially striking. A day with rainfall of 150mm causes a 4.6% increase in 5-week mortality for the youngest age group, but only a 1.5% increase for the middle age group. This finding is in sharp contrast to the temperature-mortality literature, which generally finds that impacts are concentrated on

[37] Although the middle age group exhibits larger mortality increases from the most extreme rainfall days (e.g., 300 mm) than does the oldest age group, the oldest age group exhibits larger increases on days with lower rainfall amount (e.g., < 150 mm), which occur more frequently.

the oldest age groups (Heutel et al., 2021; Carleton et al., 2022).[38] However, it is consistent with the fact that flood-related morbidities such as diarrhoea are known to strongly affect children (Kosek et al., 2003; Escobar Carias et al., 2022). Our results by age group highlight the welfare importance of investigating the impacts of rainfall shocks, since shocks that cause deaths in younger people imply a bigger welfare effect than shocks that primarily affect older populations.

In Figure 6 Panel B, we show results from a model which allows the impact of rainfall to differ by gender. We find a stronger response for women than for men, which could be due to physiological factors or social factors (e.g., more resources allocated to males in the event of a weather-induced shock) (Fruttero et al., 2023). While we cannot definitively distinguish between these two reasons, in Appendix E.2 we show results from a model in which we allow for a different response by gender and age category. The stronger female response function is driven by the middle and oldest age groups, suggesting that social factors could play a role. For the youngest age group however, boys are more affected than girls, which is consistent with the hypothesis of male biological fragility (Drevenstedt et al., 2008).

4.3 Heterogeneity in the mortality-rainfall response function across slum and non-slum populations

Next, we explore the role of socioeconomic vulnerability in determining the mortality-rainfall relationship. To do so, we create separate time series for each PIN-code for deaths in slum and non-slum addresses. We then estimate Equation (3), allowing the impacts of rainfall to vary for slum and non-slum populations.

[38] An exception is Helo Sarmiento (2023), which finds that excess mortality from hot days in Colombia is mainly driven by infectious diseases and respiratory illnesses affecting children aged 0-9.

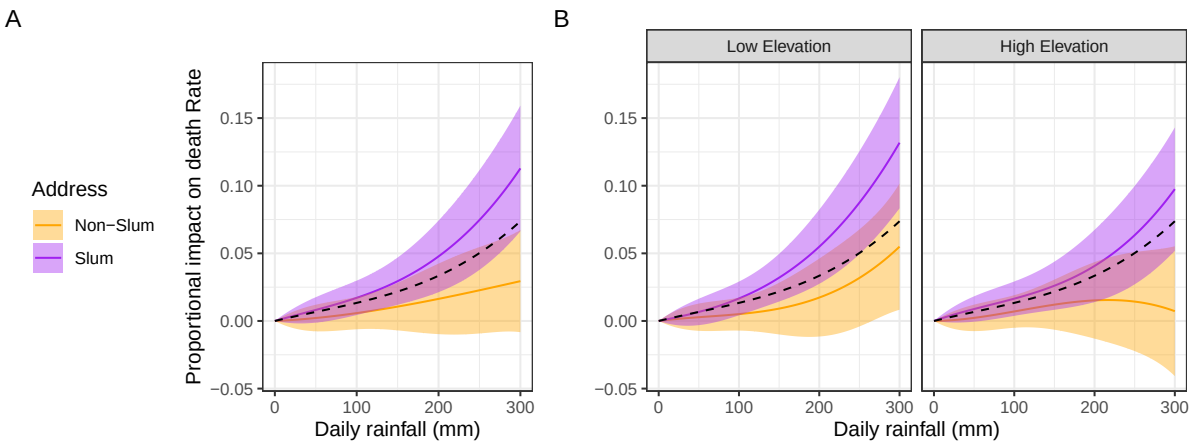

Figure 7: Percent change in 5-week mortality due to daily rainfall (slum vs. non-slum residents).

Separate mortality-rainfall response functions are estimated for slum and non-slum residents (Panel A) and for slum and non-slum residents in low- and high-elevation PIN codes (Panel B). Estimates are from a model that uses a third-order polynomial in daily rainfall for days in the contemporaneous week and 4 lagged weeks. Low (high) elevation PIN codes are defined as those whose average elevation is below (above) the median across all PIN codes. Points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. In each plot, the average response function from Figure 5 Panel A is reproduced as a dashed black line for reference.

In Figure 7, Panel A, we show estimates of the average impact of daily rainfall on mortality for slum and non-slum populations. The mortality impacts of rainfall shocks are much bigger for residents of slums than non-slums. The unequal health burden of extreme rainfall been documented qualitatively before (Patankar, 2015; WRI-India, 2022), but to our knowledge these are the first empirical, causal estimates documenting the within-city disparity of rainfall-driven mortality impacts.

In Figure 7, Panel B, we investigate whether the heterogeneity in Panel A can be explained by across-PIN code sorting with respect to elevation. It could be the case that slums are located in PIN codes that are low elevation and therefore more at risk of rainfall impacts.[39] In Panel B, we do find that low elevation PIN codes (i.e., those whose mean elevation is below the median across PIN codes) have a larger mortality response to rainfall. However, even within each elevation group, we find a substantially larger response for slum populations than non-slum populations.[40,41]

[39] In Appendix E.4 we show that the mortality-rainfall response function is much stronger in low elevation PIN codes than in high elevation PIN codes, using multiple definitions of low and high elevation.

[40] We reject a test for equality of coefficients across slum and non-slum populations, conditional on each elevation category at the 1% level in both low and high elevation PIN codes.

[41] Of course, this does not rule out that the heterogeneity is caused by sorting with respect to flood risk

From Panel B of Figure 7, we can see that impacts of rainfall are relatively small, and statistically insignificant across the full range of 0-300mm for non-slum dwellers in high elevation areas. Thus, similar to Lindersson et al. (2023), we find that accounting for inequality is a crucial factor in studies of urban flooding. Our results suggest that it is possible for some populations to be shielded entirely from the mortality impacts of rainfall, and that the impacts of urban flooding could be mitigated through adaptation investment.

In Appendix E.3 we estimate a model which allows for heterogeneity by age groups within slum and non-slum populations. Of our three categories, we see the biggest differences in the slum and non-slum response function for the middle age category, but also that slum residents are strongly affected by rainfall across all age categories. This specification shows that the difference between the slum and non-slum response function observed in Figure 7 is not driven by differences in the age profiles of slum and non-slum populations. Nevertheless, it should be noted that within the estimating sample, under-5 mortality occurs disproportionately among slum residents (Appendix Table E.1).

4.4 Evidence that urban flooding is a key mechanism for rainfall impacts

So far, we have shown that rainfall has a large and highly unequal mortality impacts. Next, we show that the mechanism for these impacts is likely through urban flooding. Establishing this mechanism is important, since urban flooding can be controlled by investment in drainage infrastructure, while rainfall is exogenous to local policy interventions.

Our results thus far already suggest that flooding is part of the mechanism through which rainfall shocks affect mortality. In Figure 7, we can see that low elevation areas face greater impacts from rainfall shocks, and this is true for both slum and non-slum populations. We illustrate the elevation heterogeneity further in Appendix E.4, using alternative measures of elevation.

Another way to understand the role of flooding is to examine the impact of hourly rainfall shocks on mortality. Figure 8 presents the results from estimating the causal response function from Equation (4), using a third-order polynomial in hourly rainfall for days in the contemporaneous week and 4 lagged weeks (Equation (5)).

within PIN codes. To investigate this possibility, we construct a cross sectional $100m^2$ pixel level dataset, which records each pixel's elevation and whether the BMC classified it as a slum area (Tandel et al., 2022). We do not find evidence that slums are located in low elevation areas within a PIN code (further details are in Appendix C.2).

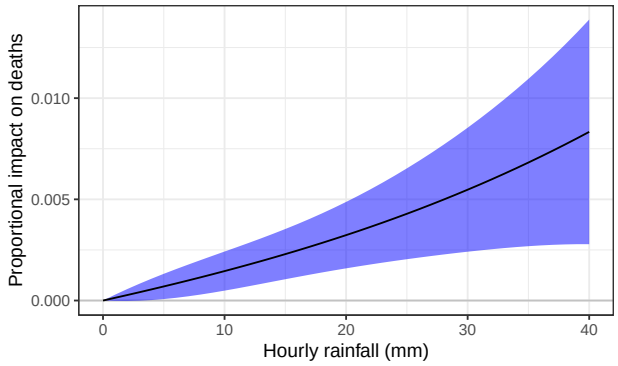

Figure 8: Percent change in 5-week mortality due to hourly rainfall.
Mortality-rainfall response function is estimated with a model that uses a third-order polynomial in hourly rainfall for hours in the contemporaneous week and 4 lagged weeks. Points along each curve represent the percent change in 5-week mortality due to a single hour with the total rainfall value shown on the x-axis, relative to an hour with zero rainfall. Shaded areas indicate 95% confidence intervals.

Figure 8 shows that on average across our sample, mortality increases sharply with hourly rainfall. Our results suggest that a single hour with 30 mm of rainfall implies an increase in 5-week mortality of over 0.5%.[42]

We further examine whether the effect of hourly rainfall on mortality is exacerbated by tidal patterns.[43] Large parts of Mumbai can be flooded when even a moderate intensity of rainfall coincides with a high tide, as the drainage network becomes overwhelmed (Zope et al., 2015).[44] If urban flooding is a key part of the mechanism underpinning the mortality-rainfall response function, then we should expect to see that the mortality impacts of rainfall are higher during high-tide periods.

Figure 9 presents the results from estimating Equation (6). We present the response function evaluated at three different values of the hourly tide, corresponding to quantiles of the in-sample distribution of hourly tide values. When the hourly tide is close to its in sample median, Model 6 gives similar results to the un-interacted model presented in Figure 8. However, when tide is high, the impacts of hourly rainfall are elevated, likely due to the increased pressure on Mumbai's ageing drainage system.

[42]The results from the hourly specification are qualitatively similar to those from the daily specification presented in Figure 5. For instance, 5 hours of 30mm rainfall would imply a mortality increase of roughly 2.5% ($5 \times 0.5\%$) in our hourly model. In the daily model, a day with 150mm of rainfall causes a mortality increase of 2.2%.

[43]Moore (2024) uses annual maximum tide height and rainfall as explanatory variables in a reduced-form empirical specification on flooding damages. However, no study to our knowledge exploits the intersection of rainfall and tide height at a subdaily level.

[44]The combination of heavy rainfall and high tide was a key factor in the city's catastrophic July 26, 2005, floods (Zope et al., 2015).

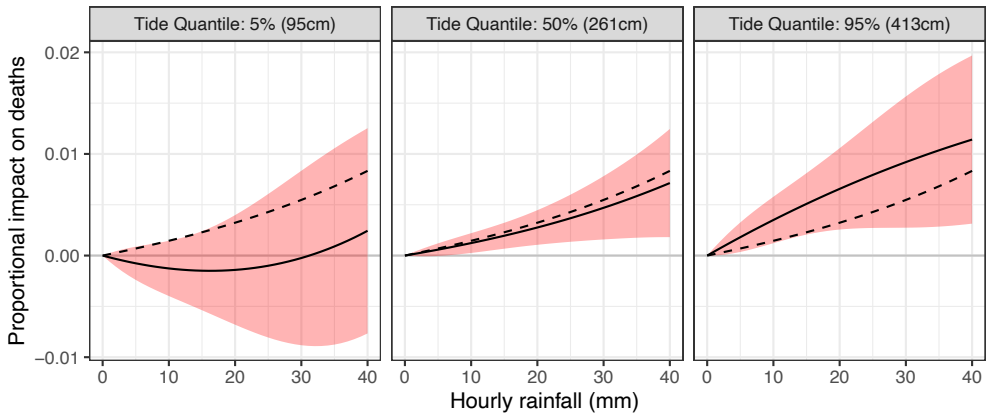

Figure 9: Percent change in 5-week mortality due to hourly rainfall, linear hourly tide interaction.

Mortality-rainfall response function is estimated with a model that uses a third-order polynomial in hourly rainfall for hours in the contemporaneous week and 4 lagged weeks. Points along each curve represent the percent change in 5-week mortality due to a single hour with the total rainfall value shown on the x-axis, relative to an hour with zero rainfall. Each panel represents the estimates from Equation 6, evaluated for the tide level at the quantile of the in-sample hourly tide distribution indicated in the facet titles. Shaded areas indicate 95% confidence intervals. In each plot, the hourly response function (without tide interaction) from Figure 8 is reproduced as a dashed black line for reference.

In Appendix E.5, we consider alternative specifications that investigate the role of tide in determining the mortality-rainfall response function. First, in Appendix E.5.1 we consider a model that allows the effect of rainfall to vary discretely based on whether the hourly tide level is above or below 450cm. In this model we find extremely large, albeit imprecisely estimated impacts for rainfall that happens during these extremely high tide levels.[45] Second, in Appendix E.5.2 we consider a model that interacts the daily rainfall response function with the daily average tide. We find that the mortality-rainfall response function becomes sharply steeper at higher daily average tide levels (Appendix Figure E.6). Together, these specifications provide strong evidence that urban flooding is a key mechanism through which rainfall affects mortality.

5 The mortality burden of rainfall shocks

The mortality-rainfall responses that we estimate in the previous section can be used to calculate the number of deaths in Mumbai that can be traced to idiosyncratic rainfall shocks

[45] While this model has the advantage of allowing the high tide response function to vary less parametrically than the model specified in Equation 6, it suffers from the fact that only 2% of the rainfall in our sample happens during periods when the tide is above 450cm.

during our estimation period.

Under our simplest model, without heterogeneous responses (Equation (1)) the log of expected deaths given a rainfall vector $\boldsymbol{R}_{i,w,y}$ is $f(\boldsymbol{R}_{i,w,y}) + \alpha_{i,y} + \lambda_{w,y}$. These deaths consist of those due to idiosyncratic rainfall shocks, and those due to predictable factors in a given PIN code and date, including due to seasonal weather patterns. We characterise the number of deaths that are not caused by idiosyncratic rainfall by evaluating Equation (1) using the fixed effects only; the log of expected deaths that are not caused by rainfall shocks is thus $\alpha_{i,y} + \lambda_{w,y}$. After converting logs to levels, we estimate the expected number of deaths caused by rainfall shocks, $r_{i,w,y}$, by subtracting total deaths from those not caused by rainfall shocks:

$$r_{i,w,y} = exp(\alpha_{i,y} + \lambda_{w,y}) \left(exp(f(\boldsymbol{R}_{i,w,y}) - 1\right) \tag{7}$$

We use our empirical estimates of all parameters in Equation (7) to estimate this object. Summing the estimate, $\hat{r}_{i,w,y}$, over the rainfall distribution in all PIN codes and weeks in our sample gives us an estimate of the total in-sample mortality burden of rainfall shocks, which can be expressed as a proportion of total observed in-sample deaths.[46] This exercise can be similarly implemented for other models estimated in the previous section, including those with heterogeneous mortality-rainfall responses by age group, slum residence, and those estimated using hourly rainfall. For instance, in the case of heterogeneity by age group, we calculate $\hat{r}_{a,i,w,y}$, i.e. the expected number of deaths in age category a, PIN code i, week w, and year y that were caused by rainfall shocks. Summing $\hat{r}_{a,i,w,y}$ over the rainfall distribution in all PIN codes and weeks in our sample gives us an estimate of the total in-sample mortality burden of rainfall for age category a, which can be expressed as a proportion of total observed in-sample deaths in age category a.[47]

Tables 4 and 5 show estimates of the proportion of deaths caused by rainfall shocks for monsoon season months (June to September) that are present in our estimating sample.[48]

[46] To quantify statistical uncertainty in this estimate, we take 1000 draws from a parametric bootstrap to draw from the parameter uncertainty of the $\hat{\beta}$ estimates that make up $\hat{g}$. We treat the fixed effects as known, as we want to compare results to a baseline, rather than consider uncertainty in our baseline. More details of the computational procedure are listed in Appendix F.1.

[47] In this case, the total in-sample mortality burden of rainfall (across all age categories) is thus calculated by summing $\hat{r}_{a,i,w,y}$ across all age categories over the rainfall distribution in all PIN codes and weeks in our sample, and expressing this as a proportion of total observed in-sample deaths. This proportion is a weighted average of the proportions calculated for the age categories, with weights equal to each age category's share of total observed in-sample deaths.

[48] Note that rainfall measurements for certain weeks in certain years are unavailable in the AWS data. To construct $\hat{r}_{i,w,y}$ we require rainfall measurements for the contemporaneous week and 4 lagged weeks, one

Each row of Table 4 displays estimates from a particular model, starting with a homogeneous mortality-rainfall response (Equation (1)) in the first row, and then proceeding to models with heterogeneous mortality-rainfall responses (Equation (3)) by age group and slum residence in the second and third rows respectively. Table 5 displays estimates from a model estimated with hourly rainfall (Equation (4)), and a model with heterogeneous mortality-rainfall responses for rainfall during low and high tide hours (Equation (6)).

		Age category			Residence	
Response	Total	0-5	5-65	65+	Non-slum	Slum
Fig 5	0.089 (0.016, 0.17)					
Fig 6	0.078 (0.0035, 0.16)	0.19 (-0.052, 0.54)	0.0063 (-0.096, 0.12)	0.13 (0.034, 0.23)		
Fig 7	0.075 (-0.00023, 0.16)				0.032 (-0.081, 0.15)	0.11 (0.0069, 0.24)

Table 4: Proportion of monsoon season deaths caused by rainfall shocks.
Table shows the proportion of total deaths and age- and residence-wise deaths during June to September that we attribute to rainfall shocks. The first row displays estimates from a model with a homogeneous mortality-rainfall response. The second and third rows respectively display estimates from models with heterogeneous mortality-rainfall responses by age and slum residence. Point estimates are calculated based on Equation (7); 95% confidence intervals, calculated with a parametric bootstrap described in Appendix F.1, are in parentheses.

Our results imply that deaths caused by rainfall shocks account for a substantial proportion of Mumbai's monsoon season deaths. In an average monsoon season, between 7% and 10% of total deaths can be traced to rainfall shocks, depending on the model. This implies that idiosyncratic rainfall during Mumbai's monsoon season kills approximately 2200 to 3000 people per year.[49] The percent of overall monsoon season deaths that we trace to rainfall shocks is comparable to the percent of the city's deaths from all cancers.[50] In Appendix F.2 we show these results are qualitatively robust under a variety of specifications, including those with more saturated fixed effects.

Models with heterogeneous mortality-rainfall responses reveal that the total mortality burden of rainfall shocks is borne unequally across different age groups and socioeconomic groups. We find that around 20% of June-September deaths for below 5 year olds, and 13% of June-September deaths for above 65 year olds, in our estimating sample can be traced to rainfall shocks, but only a negligible proportion of 5-65 year old deaths in these

or more of which may not be available, thus limiting our estimating sample.

[49] About 90,000 people die per year in Mumbai, with slightly over 1/3 of these dying during the monsoon season.

[50] In the year 2013, 7.3% of the city's total deaths were caused by cancer (Praja Foundation, 2020).

months can be traced to rainfall shocks (Table 4, Row 2).[51] Moreover, the mortality burden is disproportionately borne by slum residents, with 11% of June-September deaths at slum addresses traced to rainfall shocks compared to around 3% for non-slum addresses (Table 4, Row 3).[52]

Table 5 also reveals that the mortality due to rainfall shocks disproportionately occurs during periods of high tide as regarded by the BMC and NGOs (i.e., tide $\geq$ 450 cm). Although only 2% of rainfall over the sample period takes place during these high tide periods, these periods account for 6% of the mortality due to rainfall shocks.[53] In Appendix E.5.1, we conduct an attribution exercise using a model in which the mortality-rainfall response function is allowed to vary discretely depending on whether the tide is above or below 450cm, and find that 19% of the overall mortality due to rainfall shocks comes from high tide periods. This suggests that the effects of tide on rainfall-driven mortality are nonlinear and are likely underestimated by the linear interaction model specified in Equation 6.

[51] While we do not observe the cause of death in our data, the 20% share of under-5 deaths that we trace to rainfall is roughly similar to the share of under-5 deaths caused by diarrhoea in developing countries globally (Kosek et al., 2003).

[52] In Appendix E.3, we show the proportion of deaths caused by rainfall shocks when estimating heterogeneous mortality-rainfall responses for age $\times$ slum residence groupings (Appendix Table E.2). Under this model, roughly 7% of total monsoon season deaths in our estimating sample can be traced to rainfall shocks. It should be noted that a disproportionate share of under-5 deaths due to rainfall shocks come from slums, due to the fact that a disproportionate of share of total under-5 deaths occur in slums (Appendix Table E.1).

[53] Under the Poisson model in Equation (6), the number of total deaths due to rainfall shocks is not additively separable into deaths due to rainfall shocks during high tide and low tide. Hence the sum of values for the proportion of deaths due rainfall shocks during low tide and high tide do not exactly equal the total proportion of deaths due to rainfall shocks (Table E.3, Row 2).

Model	Total	Low Tide Rainfall	High Tide Rainfall
Hourly rainfall (Fig 8)	0.092 (0.024, 0.17)		
Linear tide interaction (Fig 9)	0.094 (0.024, 0.18)	0.089 (0.019, 0.17)	0.0053 (0.00072, 0.0094)
Proportion of rainfall	1	0.98	0.02
Proportion of attributed deaths	1	0.94	0.06

Table 5: Proportion of monsoon season deaths caused by rainfall shocks: hourly rainfall models.

The first row shows the proportion of total deaths during June to September that we attribute to rainfall shocks, derived from a model with hourly rainfall (Equation (4)). The second row shows the proportion of total deaths during June to September that we attribute to rainfall shocks during hours with high tide ($\geq$ 450 cm) and low tide ($<$ 450 cm) and overall, derived from the model specified in Equation (6). All point estimates are calculated based on Equation (7), and 95% confidence intervals (in parentheses) are calculated with a parametric bootstrap described in Appendix F.1. The third row summarises the proportion of rainfall in the sample that occurs during hours in which the average tide value is low ($<$ 450 cm) or high ($\geq$ 450 cm). The fourth row states the proportion of attributed deaths in the second row that are caused by rainfall shocks during low tide or high tide.

While estimates of r provide us with the number of deaths caused by rainfall shocks, there is also another way to interpret r. If there existed adaptive measures that could entirely eliminate excess mortality from idiosyncratic rainfall shocks (i.e., make $f(\boldsymbol{R}) = 0$),[54] then we can understand r as the number of lives that could be saved from investing in these adaptive measures. Examples of such measures may include upgrading and improving the maintenance of drainage systems. While our analysis cannot directly measure the effects of specific adaptive measures on mitigating the mortality-rainfall response, our results do suggest that there exist communities within Mumbai's for whom $f(\boldsymbol{R})$ is effectively zero. For example, Figure 7 shows roughly zero mortality response to even high levels of rainfall for non-slum residents in high elevation areas. Thus, we can think of r as comparing the excess mortality caused by rainfall shocks given currently undertaken adaptive measures, versus the number of deaths that would be caused by the same rainfall shocks under a scenario of full adaptation. Even short of full adaptation that entirely eliminates excess mortality from rainfall shocks, the results in Figure 9, Table 5 and Table E.3 suggest that such excess mortality can be substantially reduced by focused adaptive measures targeting the relatively few days where the tide reaches an extreme height (e.g., $\geq$ 450 cm). It should be noted however that any such interpretations are suggestive rather than definitive, as we cannot estimate mortality-rainfall responses with and without spe-

[54] In terms of our linear-in-parameters modelling of the mortality-rainfall response function, making $f(\boldsymbol{R}) = 0$ is equivalent to making $\beta_{l,k} = 0$ for every lag l, and polynomial power k (Equation (2)).

cific adaptive measures.

Applying an India-specific estimate of the value of a statistical life of US$ 1 million,[55] our results imply that there are potential welfare gains of around US$3 billion per year from eliminating the approximately 3,000 annual deaths from rainfall shocks. Of course, this estimate only considers the mortality costs of rainfall shocks. The overall welfare impacts would consider many other channels through which rainfall shocks and flooding affect welfare, including morbidity and economic impacts. However, the annual mortality costs of routine monsoon rainfall shocks in Mumbai are likely a substantial component of total costs, and are comparable in magnitude to direct economic damages from catastrophic flooding events. For instance, the record-setting July 26, 2005 Mumbai floods were estimated to have caused US$2 billion in direct economic damages (Ranger et al., 2011).

Our results point to multiple ways in which climate change may increase the mortality burden of rainfall. Most directly, recent evidence has suggested that climate change could increase the intensity of Asian monsoon rainfall (Salunke et al., 2023). In addition, our finding that high tides amplify the mortality impacts of rainfall suggests a channel through which rising sea levels may contribute to excess mortality, particularly in the absence of adaptive measures such as improved drainage systems. To get a sense of the possible magnitudes involved, consider the following thought experiment. Conservative estimates suggest that Mumbai will experience sea level rise of 0.5 m by 2100 (BMC, 2022). If we uniformly increase the tide level by 0.5 m across its full distribution, and re-calculate the expected mortality caused by rainfall, we find that mortality caused by rainfall increases by more than 50% when using the linear tide interaction model presented in Figure 9.[56] It should be noted however that a high tide is very different from a permanent sea level rise, and credibly assessing the impacts of the latter would require a model of the compound flooding physics.

6 Conclusion

This paper has revealed that excess mortality is a major cost of recurrent flooding in an important developing megacity— Mumbai, India. Using detailed data on person-level mortality combined with highly granular data on localised rainfall, we find that days with

[55] There is a limited literature that provides India specific VSL estimates. We use estimates consistent with Shanmugam (2011); Majumder and Madheswaran (2018); Sweis (2022).

[56] A similar exercise using the 450 cm tide threshold-based model described in Appendix E.5.1 suggests that expected mortality caused by rainfall would roughly double under 0.5 m of sea level rise.

high rainfall totals cause elevated mortality over the subsequent 5 weeks. The impacts are strongest for some of the most vulnerable groups in society— children under 5, women, and residents of slums. These costs, and their unequal distribution, have not been previously accounted for in analyses of the damages from extreme weather or urban flooding. Moreover, our results showing an increased mortality response to rainfall shocks during high tides indicate that future sea level rise could further endanger human lives.

We find that rainfall shocks not only cause significant excess mortality but also explain a substantial proportion of Mumbai's overall mortality. This stands in contrast to previous studies on the mortality impacts of extreme temperatures, which find no effect of temperature shocks on excess mortality in urban India (Burgess et al., 2017). Our results suggest that policy makers should not ignore the harms from extreme rainfall either today or under future climate change, and should target resources towards vulnerable groups. The large proportion of under-5 mortality explained by rainfall shocks points to large potential gains from improvements in facilities that are lacking in slum settings (e.g., adequate housing, safe water, and sanitation), particularly given that overall under-5 mortality disproportionately occurs among slum residents (Appendix E.1). Investments in safe water and sanitation have been found to drive historical declines in urban child mortality in today's developed countries (Alsan and Goldin, 2019; Anderson et al., 2022). Similar declines could potentially be achieved in today's developing urban areas.

Several caveats and limitations are worth noting, and point to directions for future research. First, we have utilised mortality and rainfall data of unprecedented granularity within an urban area, but at the same time, these data come with limitations. The mortality data do not report the cause of death, precluding us from fully understanding the exact ways in which rainfall shocks imperils human health. Moreover, the rainfall data make it difficult to capture mortality impacts that may occur over longer time horizons. In particular, the measurements of local rainfall are not available over all days of the year, preventing us from estimating models that include even more temporal lags of rainfall. Accounting for these longer term effects may be necessary given the extended duration certain monsoon-related diseases can take to develop.[57] Additionally, our fixed effects remove the effects of seasonally predictable shocks. Thus our estimates can be interpreted as a lower bound on the total mortality caused by rainfall shocks.

Second, our findings that the mortality impacts of rainfall are stronger in low elevation

[57]Malaria is one such example, which can have an incubation of 30 days before the onset of symptoms (Liang and Messenger, 2018). Thus, the mortality impacts of a malaria outbreak associated with a flooding event might not be observed until outside of the 5 week window used in our study.

PIN codes and during periods of high tide suggest that flooding is a key channel for these impacts. However, we do not directly observe flood depth or duration across locations following rainfall events, and thus cannot directly estimate the response of mortality to flood depth and duration. Observing flooding directly would make it possible to explore the pathways through which rainfall shocks propagate across space and harm human health, and also shed further light on the value of improvements in drainage infrastructure.

Finally, while this paper's findings highlight large potential benefits of protective investments that break the link between rainfall shocks and excess mortality, especially for the worst affected groups, future research should quantify the benefits of specific investments in urban settings. These may include improvements to housing, water, sanitation, and infrastructure, protection of natural drainage sources such as wetlands, as well as improvements to and proper maintenance of man-made drainage systems.[58] As we enter an era where most of the urban population growth is occurring in the developing world and much of that in slums, understanding the returns to these investments is a critical area for future research.

[58] In the case of Mumbai and other developing megacities, drainage systems are antiquated and are often clogged with solid waste (Ranger et al., 2011; De Sherbinin et al., 2012; Patankar, 2015; Singh, 2018). Ongoing urbanisation has further exacerbated flood risk, as natural drainage sources such as mangroves are lost to construction (Singh, 2018).

References

Alsan, Marcella and Claudia Goldin, "Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920," *Journal of Political Economy*, 2019, *127* (2), 586–638. Publisher: The University of Chicago Press Chicago, IL.

Anderson, D Mark, Kerwin Kofi Charles, and Daniel I Rees, "Reexamining the contribution of public health efforts to the decline in urban mortality," *American Economic Journal: Applied Economics*, 2022, *14* (2), 126–157. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

Ashraf, Nava, Edward Glaeser, Abraham Holland, and Bryce Millett Steinberg, "Water, health and wealth: The impact of piped water outages on disease prevalence and financial transactions in zambia," *Economica*, 2021, *88* (351), 755–781. Publisher: Wiley Online Library.

Auffhammer, Maximilian, Solomon M. Hsiang, Wolfram Schlenker, and Adam Sobel, "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change," *Review of Environmental Economics and Policy*, July 2013, *7* (2), 181–198. Publisher: The University of Chicago Press.

Banzhaf, Spencer, Lala Ma, and Christopher Timmins, "Environmental justice: The economics of race, place, and pollution," *Journal of Economic Perspectives*, 2019, *33* (1), 185–208. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418.

Bhalotra, Sonia R, Alberto Diaz-Cayeros, Grant Miller, Alfonso Miranda, and Atheendar S Venkataramani, "Urban water disinfection and mortality decline in lower-income countries," *American Economic Journal: Economic Policy*, 2021, *13* (4), 490–520. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

BMC, "Mumbai Climate Action Plan," 2022.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis, "Cities in the developing world," *Annual Review of Economics*, 2020, *12* (Volume 12, 2020), 273–297.

Burgess, Robin, Olivier Deschenes, Dave Donaldson, and Michael Greenstone, "Weather, climate change and death in India," *University of Chicago*, 2017.

Cain, Lucas, Danae Hernandez-Cortes, Christopher Timmins, and Paige Weber, "Recent findings and methodologies in economics research in environmental justice," *Review of Environmental Economics and Policy*, 2024, *18* (1), 000–000. Publisher: The University of Chicago Press Chicago, IL.

Carias, Michelle S Escobar, David W Johnston, Rachel Knott, and Rohan Sweeney, "Flood disasters and health among the urban poor," *Health Economics*, 2022, *31* (9), 2072–2089. Publisher: Wiley Online Library.

Carleton, Tamma, Amir Jina, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, Andrew Hultgren, Robert E Kopp, Kelly E McCusker, Ishan Nath, James Rising, Ashwin Rode, Hee Kwon Seo, Arvid Viaene, Jiacan Yuan, and Alice Tianbo Zhang, "Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits*," *The Quarterly Journal of Economics*, November 2022, *137* (4), 2037–2105.

Cattaneo, Matias D, Sebastian Galiani, Paul J Gertler, Sebastian Martinez, and Rocio Titiunik, "Housing, health, and happiness," *American Economic Journal: Economic Policy*, 2009, *1* (1), 75–105. Publisher: American Economic Association.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano, "Catastrophic natural disasters and economic growth," *Review of Economics and Statistics*, 2013, *95* (5), 1549–1561. Publisher: The MIT Press.

Centre for Research on the Epidemiology of Disasters, "EM-DAT: The CRED international disaster database," 2023. Place: Brussels.

Chakma, Tridevi, Jonathan Colmer, and John Voorheis, "The causes and consequences of urban heat islands," Technical Report 2024.

Chen, Jiafeng and Jonathan Roth, "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, 2024, *139* (2), 891–936. Publisher: President and Fellows of Harvard College.

Chen, Joyce J, Valerie Mueller, Yuanyuan Jia, and Steven Kuo-Hsin Tseng, "Validating migration responses to flooding using satellite and vital registration data," *American Economic Review*, 2017, *107* (5), 441–445. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Cohen, François and Antoine Dechezleprêtre, "Mortality, Temperature, and Public Health Provision: Evidence from Mexico," *American Economic Journal: Economic Policy*, May 2022, *14* (2), 161–192.

Collier, Paul and Anthony Venables, "Urbanization in developing economies: the assessment," *Oxford Review of Economic Policy*, 2017, *33* (3), 355–372. Publisher: Oxford University Press and Oxford Review of Economic Policy Limited.

Cutler, David and Grant Miller, "The role of public health improvements in health advances: the twentieth-century United States," *Demography*, 2005, *42* (1), 1–22. Publisher: Springer.

Damania, R., S. Desbureaux, and E. Zaveri, "Does rainfall matter for economic growth? Evidence from global sub-national data (1990–2014)," *Journal of Environmental Economics and Management*, July 2020, *102*, 102335.

Deschênes, Olivier and Enrico Moretti, "Extreme Weather Events, Mortality, and Migration," *The Review of Economics and Statistics*, 2009, *91* (4), 659–681. Publisher: MIT Press.

_ **and Michael Greenstone**, "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US," *American Economic Journal: Applied Economics*, October 2011, *3* (4), 152–185.

Desmet, Klaus, Robert E. Kopp, Scott A. Kulp, Dávid Krisztián Nagy, Michael Oppenheimer, Esteban Rossi-Hansberg, and Benjamin H. Strauss, "Evaluating the Economic Cost of Coastal Flooding," *American Economic Journal: Macroeconomics*, April 2021, *13* (2), 444–486.

Donaldson, Dave and Adam Storeygard, "The view from above: Applications of satellite data in economics," *Journal of Economic Perspectives*, 2016, *30* (4), 171–198. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418.

Drevenstedt, Greg L, Eileen M Crimmins, Sarinnapha Vasunilashorn, and Caleb E Finch, "The rise and fall of excess male infant mortality," *Proceedings of the National academy of Sciences*, 2008, *105* (13), 5016–5021. Publisher: National Acad Sciences.

Elliott, Robert JR, Eric Strobl, and Puyang Sun, "The local impact of typhoons on economic activity in China: A view from outer space," *Journal of Urban Economics*, 2015, *88*, 50–66. Publisher: Elsevier.

Fruttero, Anna, Daniel Halim, Chiara Broccolini, Bernardo Coelho, Horace Gninafon, and Noël Muller, *Gendered Impacts of Climate Change: Evidence from Weather Shocks* Policy Research Working Papers, The World Bank, May 2023.

Galiani, Sebastián, Paul J. Gertler, Raimundo Undurraga, Ryan Cooper, Sebastián Martínez, and Adam Ross, "Shelter from the storm: Upgrading housing infrastructure in Latin American slums," *Journal of Urban Economics*, March 2017, *98*, 187–213.

Gamper-Rabindran, Shanti, Shakeeb Khan, and Christopher Timmins, "The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach," *Journal of Development Economics*, 2010, *92* (2), 188–200. Publisher: Elsevier.

Gandhi, Sahil, Matthew E. Kahn, Rajat Kochhar, Somik Lall, and Vaidehi Tandel, "Adapting to Flood Risk: Evidence from a Panel of Global Cities," June 2022.

Gillingham, Kenneth and Pei Huang, "Racial disparities in the health effects from air pollution: Evidence from ports," Technical Report, National Bureau of Economic Research 2021.

Guan, Mingfu, Qiuhua Liang, and Jingming Hou, "Smart approaches to predict urban flooding: Current advances and challenges," *Frontiers in Earth Science*, 2021, *9*, 681751. Publisher: Frontiers Media SA.

Guiteras, Raymond, Amir Jina, and A. Mushfiq Mobarak, "Satellites, Self-reports, and Submersion: Exposure to Floods in Bangladesh," *The American Economic Review*, 2015, *105* (5), 232–236. Publisher: American Economic Association.

Henderson, J. Vernon and Matthew A. Turner, "Urbanization in the Developing World: Too Early or Too Slow?," *Journal of Economic Perspectives*, August 2020, *34* (3), 150–173.

Heutel, Garth, Nolan H. Miller, and David Molitor, "Adaptation and the Mortality Effects of Temperature across U.S. Climate Regions," *The Review of Economics and Statistics*, September 2021, *103* (4), 740–753.

Holtermann, Linus, "Precipitation anomalies, economic production, and the role of "first-nature" and "second-nature" geographies: A disaggregated analysis in high-income countries," *Global Environmental Change*, November 2020, *65*, 102167.

Hsiang, Solomon, "Climate Econometrics," *Annual Review of Resource Economics*, October 2016, *8* (1), 43–75. Publisher: Annual Reviews.

Hsiang, Solomon M and Amir S Jina, "The causal effect of environmental catastrophe on long-run economic growth: Evidence from 6,700 cyclones," Technical Report, National Bureau of Economic Research 2014.

Hsiao, Allan, "Sea Level Rise and Urban Adaptation in Jakarta," 2023.

Islam, Md Tazmul and Qingmin Meng, "An exploratory study of Sentinel-1 SAR for rapid urban flood mapping on Google Earth Engine," *International Journal of Applied Earth Observation and Geoinformation*, 2022, *113*, 103002. Publisher: Elsevier.

Jha, Abhas K, Robin Bloch, and Jessica Lamond, *Cities and flooding: a guide to integrated urban flood risk management for the 21st century*, World Bank Publications, 2012.

Kahn, Matthew E, "The death toll from natural disasters: the role of income, geography, and institutions," *Review of economics and statistics*, 2005, *87* (2), 271–284. Publisher: MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046, USA journals ....

Kocornik-Mina, Adriana, Thomas K. J. McDermott, Guy Michaels, and Ferdinand Rauch, "Flooded Cities," *American Economic Journal: Applied Economics*, April 2020, *12* (2), 35–66.

Kolstad, Charles D. and Frances C. Moore, "Estimating the Economic Impacts of Climate Change Using Weather Observations," *Review of Environmental Economics and Policy*, January 2020, *14* (1), 1–24. Publisher: The University of Chicago Press.

Kosek, Margaret, Caryn Bern, and Richard L Guerrant, "The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000," *Bulletin of the world health organization*, 2003, *81*, 197–204. Publisher: SciELO Public Health.

Kotz, Maximilian, Anders Levermann, and Leonie Wenz, "The effect of rainfall changes on economic production," *Nature*, January 2022, *601* (7892), 223–227. Number: 7892 Publisher: Nature Publishing Group.

Kuddus, Md Abdul, Elizabeth Tynan, and Emma McBryde, "Urbanization: a problem for the rich and the poor?," *Public Health Reviews*, January 2020, *41* (1), 1.

Liang, Stephen Y. and Nicole Messenger, "Infectious diseases after hydrologic disasters," *Emergency medicine clinics of North America*, November 2018, *36* (4), 835–851.

Lindersson, Sara, Elena Raffetti, Maria Rusca, Luigia Brandimarte, Johanna Mård, and Giuliano Di Baldassarre, "The wider the gap between rich and poor the higher the

flood mortality," *Nature Sustainability*, August 2023, *6* (8), 995–1005. Number: 8 Publisher: Nature Publishing Group.

Maccini, Sharon and Dean Yang, "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *The American Economic Review*, 2009, *99* (3), 1006–1026. Publisher: American Economic Association.

Majumder, Agamoni and S. Madheswaran, "Value of statistical life in India: A hedonic wage approach," Working Paper 407, Institute for Social and Economic Change, Bangalore 2018.

Marx, Benjamin, Thomas Stoker, and Tavneet Suri, "The economics of slums in the developing world," *Journal of Economic perspectives*, 2013, *27* (4), 187–210. Publisher: American Economic Association.

Mittra, Indraneel, Gauravi A. Mishra, Rajesh P. Dikshit, Subhadra Gupta, Vasundhara Y. Kulkarni, Heena Kauser A. Shaikh, Surendra S. Shastri, Rohini Hawaldar, Sudeep Gupta, C. S. Pramesh, and Rajendra A. Badwe, "Effect of screening by clinical breast examination on breast cancer incidence and mortality after 20 years: prospective, cluster randomised controlled trial in Mumbai," *BMJ*, February 2021, *372*, n256. Publisher: British Medical Journal Publishing Group Section: Research.

Moore, Frances C, "Learning, catastrophic risk, and ambiguity in the climate change era," *NBER Chapters*, 2024. Publisher: National Bureau of Economic Research, Inc.

National Institute of Oceanography, "Study of Land Reclamation and its Influence on Mumbai's Coastline and Implications for Proposed Projects," Technical Report, Mumbai Transformation Support Unit 2016.

Patankar, Archana Mahesh, "The exposure, vulnerability, and ability to respond of poor households to recurrent floods in Mumbai," *World Bank Policy Research Working Paper*, 2015, *1* (7481).

Patel, Dev, "Floods," Technical Report 2024.

Picarelli, Nathalie, Pascal Jaupart, and Ying Chen, "Cholera in times of floods: Weather shocks and health in Dar es Salaam," *International Growth Centre Working Paper*, 2017.

Praja Foundation, "Cause of death in mumbai: 2012 to 2020," 2020. Place: Mumbai.

Ranger, Nicola, Stéphane Hallegatte, Sumana Bhattacharya, Murthy Bachu, Satya Priya, K Dhore, Farhat Rafique, P Mathur, Nicolas Naville, Fanny Henriet, and others, "An assessment of the potential impact of climate change on flood risk in Mumbai," *Climatic change*, 2011, *104*, 139–167. Publisher: Springer.

Rentschler, Jun, Paolo Avner, Mattia Marconcini, Rui Su, Emanuele Strano, Michalis Vousdoukas, and Stéphane Hallegatte, "Global evidence of rapid urban growth in flood zones since 1985," *Nature*, October 2023, *622* (7981), 87–92. Number: 7981 Publisher: Nature Publishing Group.

Russ, Jason, "Water runoff and economic activity: The impact of water supply shocks on growth," *Journal of Environmental Economics and Management*, May 2020, *101*, 102322.

Salunke, Popat, Narayan Prasad Keshri, Saroj Kanta Mishra, and S. K. Dash, "Future projections of seasonal temperature and precipitation for India," *Frontiers in Climate*, 2023, *5*.

Sarmiento, Juliana Helo, "Into the tropics: Temperature, mortality, and access to health care in Colombia," *Journal of Environmental Economics and Management*, May 2023, *119*, 102796.

Schiavina, Marcello, Thomas Kemper, Martino Pesaresi, Daniele Ehrlich, Michele Melchiorri, Pietro Florio, Alessandra Carioli, Johannes Uhl, Katarzyna Goch, Panagiotis Politis, Luca Maffenini, Pierpaolo Tommasi, and Ines Rivero, *GHSL Data Package 2023* May 2023.

Schumann, Guy, Laura Giustarini, Angelica Tarpanelli, Ben Jarihani, and Sandro Martinis, "Flood modeling and prediction using earth observation data," *Surveys in Geophysics*, 2023, *44* (5), 1553–1578. Publisher: Springer.

Shanmugam, K. R., "Discount Rate for Health Benefits and the Value of Life in India," Working Paper, Madras School of Economics, Chennai, India August 2011.

Sherbinin, Alex De, Andrew Schiller, and Alex Pulsipher, "The vulnerability of global cities to climate hazards," in "Adapting cities to climate change," Routledge, 2012, pp. 129–157.

Singh, D, "Mumbai: The 'Rain Ready'city that floods every year," *Category: Mumbai* [*Article*]. *Available online at: http://www. acclimatise. uk. com/tag/mumbai/*[*Accessed 27/11/2019*], 2018.

Singh, Jitendra, Sheeba Sekharan, Subhankar Karmakar, Subimal Ghosh, P. E. Zope, and T. I. Eldho, "Spatio-temporal analysis of sub-hourly rainfall over Mumbai, India: Is statistical forecasting futile?," *Journal of Earth System Science*, April 2017, *126*, 38. ADS Bibcode: 2017JESS..126...38S.

Sweis, Nadia J., "Revisiting the value of a statistical life: an international approach during COVID-19," *Risk Management*, 2022, *24* (3), 259–272.

Tandel, Vaidehi, Sahil Gandhi, Shaonlee Patranabis, Luís M. A. Bettencourt, and Anup Malani, "Infrastructure, enforcement, and COVID-19 in Mumbai slums: A first look," *Journal of Regional Science*, 2022, *62* (3), 645–669. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jors.12552.

Tanim, Ahad Hasan, Callum Blake McRae, Hassan Tavakol-Davani, and Erfan Goharian, "Flood detection in urban areas using satellite imagery and machine learning," *Water*, 2022, *14* (7), 1140. Publisher: MDPI.

Tripathy, Shrabani S., Hari Vittal, Subhankar Karmakar, and Subimal Ghosh, "Flood risk forecasting at weather to medium range incorporating weather model, topography, socio-economic information and land use exposure," *Advances in Water Resources*, December 2020, *146*, 103785.

Tuholske, Cascade, Kelly Caylor, Chris Funk, Andrew Verdin, Stuart Sweeney, Kathryn Grace, Pete Peterson, and Tom Evans, "Global urban population exposure to extreme heat," *Proceedings of the National Academy of Sciences*, 2021, *118* (41), e2024792118. Publisher: National Acad Sciences.

UN DESA, "World urbanization prospects: The 2018 revision," 2018.

UNISDR, "From shared risk to shared Value–The business case for disaster risk reduction," *Global Assessment Report on Disaster Risk Reduction*, 2013. Publisher: United Nations Office for Disaster Risk Reduction (UNISDR) Geneva, Switzerland.

Wang, Bin, Michela Biasutti, Michael P. Byrne, Christopher Castro, Chih-Pei Chang, Kerry Cook, Rong Fu, Alice M. Grimm, Kyung-Ja Ha, Harry Hendon, Akio Kitoh, R. Krishnan, June-Yi Lee, Jianping Li, Jian Liu, Aurel Moise, Salvatore Pascale, M. K. Roxy, Anji Seth, Chung-Hsiung Sui, Andrew Turner, Song Yang, Kyung-Sook Yun, Lixia Zhang, and Tianjun Zhou, "Monsoons Climate Change Assessment," *Bulletin of the American Meteorological Society*, January 2021, *102* (1), E1–E19. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

World Economic Forum, "These are the world's five biggest slums," 2020.

WRI-India, "Climate and Air Pollution Risks and Vulnerability Assessment for Mumbai, India," Technical Report 2022.

Zope, P. E., T. I. Eldho, and V. Jothiprakash, "Impacts of urbanization on flooding of a coastal urban catchment: a case study of Mumbai City, India," *Natural Hazards*, 2015, *75* (1), 887–908. Publisher: Springer.

Contents

A Seasonality in disease incidence

To investigate the seasonality in diseases across Mumbai, we obtain monthly data on disease cases reported at each of the medical dispensaries run by the BMC, through correspondence with a non-governmental organisation, the Praja Foundation.[59] Figure A.1 plots time series and monthly sums of these data, for four diseases that are known to be related to urban flooding. All four show clear signs of elevated incidence during monsoon months.

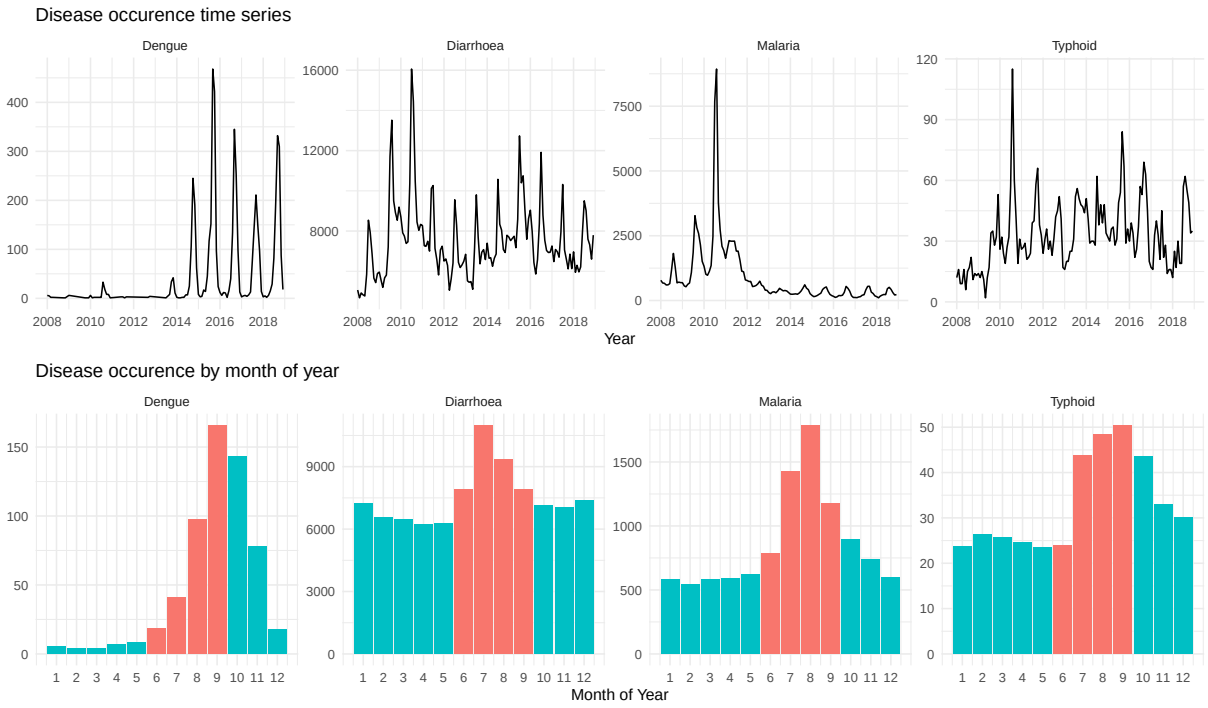

Figure A.1: Seasonality of disease incidence
Top panel shows city-wide incidence for four selected diseases at the monthly level between 2008 and 2019. Bottom panel takes monthly means across the years to illustrate seasonality. Monsoon season months are coloured in red. We can see clear evidence of elevated incidence during monsoon months across each of these disease types. Data is obtained from the Praja Foundation, who collected the raw data through a Right To Information (RTI) request.

[59] https://www.praja.org/

B Data

B.1 Location assignment in mortality data

We assign deaths to locations using one of two approaches. In the approach used in our main specification, we use the PIN code information from the addresses provided in the raw data. However, this approach requires us to throw out roughly 20% of our observations, for which we do not have a PIN code. Therefore in an alternative approach, we assign deaths to each of Mumbai's 24 municipal wards, using the ward indicated in the raw data. Sensitivity of our results to these two approaches is shown in Appendix D.8.

B.2 Slum assignment

For each address in our sample, we categorise whether it was likely to have been a slum address using the following procedure.

First, we obtain a list of slum names from the 2001 census. These data also contain the municipal ward in which each named slum is located. Next, we conduct a fuzzy match on these slum names, where we match slum names from a given ward into addresses from PIN codes that overlap with these wards. Following this, we remove a number of false positive matches, that are due to highly generic slum names in the census list.[60] We also supplement this list, based on a text analysis of the most common terms that appear in our address lists, combined with local knowledge of which of these common terms are likely to be associated with slum addresses. Specifically, these terms are *zopadpatti* (the Marathi language word for slum, including abbreviated forms *zop*, *zp*, and *z.p*), *chawl* (a type of dense of communal housing, including abbreviated form *chl*), "police" (addresses of homeless persons or residents of informal settlements may be recorded as the nearest police station), *nagar* (literally a district or quarter within a town, but in Mumbai commonly applied to slum settlements), and "transit". The word "transit" is included to capture "transit camps", which are temporary settlements for persons displaced due to construction projects or from dilapidated buildings about to be destroyed or redeveloped. In practice, however, transit camps can persist as long-term settlements.[61]

[60] Specifically, some slum names are identical to those of large Mumbai neighbourhoods that contain a variety of housing types. These names include Malad, Borivali, Ghatkopar, Kandivali, Andheri, Goregaon, Jogeshwari, Byculla, Chembur, Vile Parle, and Colaba.

[61] See for example `https://www.timesnownews.com/mirror-now/in-focus/people-living-in-transit-camps-for-30-years-redevelopment-pending-for-decades-article-93362669`.

Overall, we categorise roughly 55% of our sample as slum addresses, which corresponds well to official estimates of Mumbai's slum population.

We use two methods to further validate our approach. First, we compare the spatial distribution of the categorisation to the proportion of people categorised as living in slums in each municipal ward as per the 2011 census. We find the two measures to be strongly correlated (see Figure B.1). Second, we compare the mean age of death in the slum addresses to that in the non-slum addresses. We find that the mean age of death is much lower in our slum addresses, which is consistent with the fact that poverty and its associated health problems are much more common in slums than non-slums.

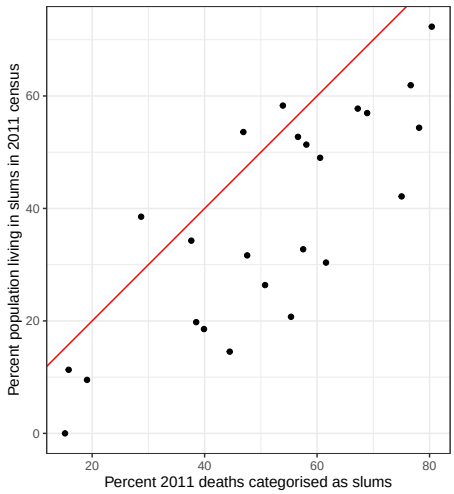

Figure B.1: Comparison of slum proportion estimates to census data
The 2011 census includes estimates of the proportion of residents of each ward that live in slums. Plot compares these census values to our estimates of the proportion of deaths in each ward that are from people who live in addresses we classify as slum addresses. The red-line shows a 45 degree line. The R^2 value for this relationship is around 0.63.

B.3 Weather data

We obtain weather data at fifteen minute intervals from the BMC for the years 2006 to 2015. These data provide us with weather observations from a large set of automatic weather stations (AWS) that were set up by Mumbai's local government following the July 2005 floods.

Figure B.2 shows the locations of the set of stations we have available to us in each year.

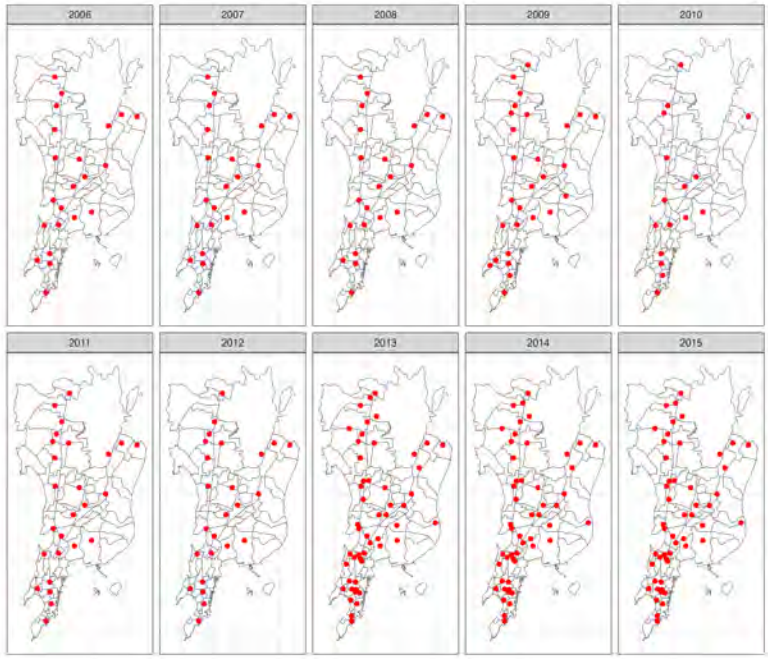

Figure B.2: Locations of AWS weather stations
Plot shows the geo-locations of the AWS weather stations available to us in each year of our estimating sample. There are 13 stations that appear in every year's sample.

The network is extremely dense, and provides good geographical coverage across the city. This is important, since Mumbai's rainfall is highly localised. However, the data required considerable processing in order to be usable for regression analysis. We next outline the steps we undertook to process these station level data, and merge them with our mortality data.

Step 1: Cleaning station names and dates

First, we match station names in our AWS data to a master list of AWS stations and their latitude and longitude information, and clean the dates which are inconsistently formatted. The output of this first step is an unbalanced panel of station level observations. Most of these observations are 15 minute observations. We have different stations and temporal coverage available in each year.

Step 2: Creating a subdaily station level panel

Second, we aggregate this sub-daily information into a balanced panel of station-15 minute period level observations for each year. During this procedure, we implement the follow-

ing data cleaning steps:

- We drop station-days in each year where there are fewer than half the stations available.
- We linearly interpolate short strings of missing data. In particular, if a station-day contains a string of 1 hour or less of rainfall, or a string of 4 hours or less of temperature, we linearly interpolate to fill in those missing values.
- For some station-periods, we have more than one observation in the raw data. For those that we cannot determine which should take precedence, we average over the multiple observations.
- We remove station-period observations in which the temperature is recorded to be lower than 0 C. These temperature values are not plausible for Mumbai, and appear correlated with inaccurate rainfall measurements.
- We replace implausible temperature and rainfall values with NAs. We define implausible values (at the 15-minute by station level) as:
  - Temperature above 42.2C, or below 10C (this is the maximum and minimum historical values in Mumbai).
  - Above 50mm of rainfall, which we deem to be implausible.
  - Temperature values that are more than 2 standard deviations outside of the monthly mean for their time of the day across all available stations.
  - We also replace the rainfall value from the Malad station for 2009-07-23 with an NA, due to idiosyncrasies[62] in the rainfall data.
- Imputation of missing station periods
  - In order to keep the set of stations consistent within a year, we next interpolate the missing station-day values using multiple imputation. We do this using the MICE R package
  - We constrain this process such that an imputed station-day must fall within the range of the other stations on that day, to ensure that our imputation doesn't cause us to create artificially extreme values.

[62]Specifically, we find that this value is very extreme, relative to both the distribution of other AWS stations on this day, and compared to our IMD data. We also did not find any news reports to suggest that there was an extremely large rainfall shock in the Malad area on this day.

- Overall, we end up imputing around 9% of the station-15 minute period temperature values, and around 5% of the rainfall values.

We conduct extensive diagnostics on these station level values, including comparisons to our purchased Indian Meteorological Department (IMD) data, in order to ensure their quality and iron out idiosyncrasies.

Step 3: Creating zonal statistics

Step 2 provided us with a panel of station-period level rainfall and temperature data, that is balanced within a year, but not across years.

Our next step is to create PIN code and ward (feature) level zonal statistics. To do this, we calculate the distance between the centroid of each feature and each weather station. We then take an inverse-distance squared weighted average of the stations to calculate each feature's temperature and rainfall values for each 15 minute period for which we have coverage.

Step 4: Collapsing to the daily level

We next collapse our subdaily data to the daily level, taking transformations of the subdaily data as appropriate. For instance, we calculate polynomials in the hourly rainfall, and calculate interactions of subdaily tide and rainfall.

Step 5: Non-linear transformations of the daily data

We now have a balanced panel of feature-day level weather variables. We take non-linear transformations at the daily level, in order to preserve important non-linearities that happen at the daily level (Hsiang, 2016). These include binning, splines, and polynomials.

Step 6: Temporal aggregation

Finally, we aggregate the daily data to the weekly or monthly level that is used in our regressions by taking a simple summation.

B.3.1 Rainfall distribution across the city

Table B.1: Variation in climate variables

Table shows the PIN code level summary statistics of the daily rainfall distribution across our estimating sample. The rainfall column shows the daily average rainfall across the estimating sample. $q99$ denotes the 99^{th} percentile of the daily distribution. Finally, the 'Days above 100mm' column denotes the number of days each pincode experiences at least 100mm of rainfall. Overall, despite the high level of spatial variation on individual days, the total rainfall and prevalence of extreme rainfall within a year is roughly uniform across the city.

Pincode	Rainfall	q99	Days > 100mm	Pincode	Rainfall	q99	Days > 100mm
400001	10.92	125.26	38	400057	11.31	127.91	35
400002	10.41	123.11	35	400058	11.48	135.99	41
400003	10.93	128.93	42	400059	11.39	130.76	47
400004	10.53	123.91	35	400060	11.46	128.37	42
400005	10.36	123.80	37	400061	11.35	121.79	41
400006	10.37	124.92	34	400063	11.40	124.86	36
400007	10.06	125.98	37	400064	10.33	126.06	35
400008	10.81	125.33	39	400065	11.43	123.29	39
400009	11.10	128.56	43	400066	11.70	118.89	39
400010	10.92	125.33	41	400067	11.13	124.94	35
400011	10.87	127.65	40	400068	12.44	141.38	48
400012	11.03	123.21	44	400069	11.48	135.10	40
400013	10.90	123.83	40	400070	12.53	144.46	48
400014	10.91	130.78	46	400071	11.70	133.55	45
400015	11.11	124.51	40	400072	11.88	134.57	46
400016	11.32	132.03	46	400074	11.66	130.64	46
400017	11.41	133.47	47	400075	12.05	135.00	47
400018	10.87	124.71	39	400076	12.03	128.51	48
400019	11.52	129.83	46	400077	12.28	143.93	47
400020	10.41	123.53	37	400078	11.55	120.65	42
400021	10.27	124.24	36	400079	12.03	134.32	46
400022	11.57	132.68	43	400080	12.92	141.04	44
400024	11.84	135.28	46	400081	12.81	133.15	46
400025	10.56	123.78	43	400082	12.74	138.01	43

400026	10.35	122.85	35	400083	12.02	125.76	47
400027	10.96	134.66	43	400084	12.33	140.85	49
400028	11.10	125.68	43	400085	11.59	127.32	44
400029	11.73	137.55	43	400086	12.16	140.09	46
400030	10.51	121.53	43	400087	11.69	125.07	45
400031	11.13	127.49	42	400088	11.80	128.49	42
400032	10.34	123.25	38	400089	12.15	138.07	46
400033	11.01	125.63	40	400091	10.99	122.63	36
400034	10.52	123.64	35	400092	11.32	120.93	36
400035	10.55	124.58	36	400093	11.48	129.18	41
400037	11.48	135.46	42	400094	11.73	129.36	45
400042	12.04	124.25	45	400095	11.20	124.94	36
400043	11.93	129.72	46	400096	11.51	125.74	41
400049	11.33	127.44	38	400097	11.34	122.44	32
400050	11.16	133.41	44	400098	11.52	138.47	45
400051	11.45	136.26	45	400099	11.74	134.33	45
400052	11.31	131.80	42	400101	11.30	123.75	36
400053	11.42	125.96	41	400102	11.47	128.11	41
400054	11.32	130.63	36	400103	11.78	131.39	42
400055	11.43	133.51	37	400104	11.34	127.53	37
400056	11.31	131.52	36	NA	NA	NA	NA

B.3.2 Comparison of AWS and IMD data

While we do not have a method of validating the spatial variation in our AWS rainfall and temperature variables, we can compare the temporal variation to that from the IMD's weather station at Mumbai's airport. Overall the temporal patterns are similar in the two datasets.

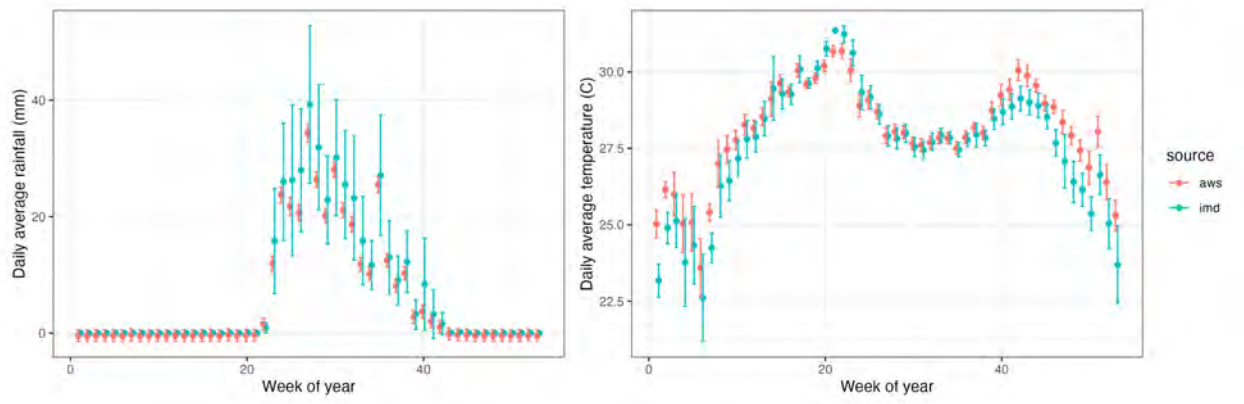

Figure B.3: AWS and IMD comparison
Plot shows average temperature and precipitation values by week of year in the AWS data used in our analysis, compared to the average temperature and precipitation values by week of year in the Indian Meteorological Department's weather station data. The two measures show similar overall patterns. AWS data are averaged across available stations, while the IMD data is from the Santa Cruz airport weather station.

B.4 Estimating sample

Year	Total deaths in raw data	Total deaths in cleaned data	Weeks in estimating sample	Total deaths in estimating sample
2006	89661	67729	30 to 52	31254
2007	91011	73917	1 to 52	73695
2008	90127	73746	1 to 39	55559
2009	90650	72870	6 to 39	47119
2010	97010	76870	27 to 43	28889
2011	89390	70177	27 to 45	26362
2012	88925	69161	27 to 34	11265
2013	90078	69389	27 to 30	5328
2014	92447	70980	27 to 43	24285
2015	91797	69718	27 to 39	17496

Table B.2: Weeks in estimating sample by year
Table shows a breakdown of the number of deaths in our raw data compared to the number of deaths in our estimating sample. Total deaths in cleaned data are lower than total deaths in raw data due to missing PIN codes in the addresses of some observations. Out of the total deaths in the cleaned data, the total deaths used in the estimating sample is determined by the availability of the AWS data, which are mainly available only during the monsoon season in each year.[64]

[64]Note, weeks are defined relative to Jan 1st in each year. Thus, we have a few days per year in a 53rd week, which we drop. This causes us to drop 176 deaths from the cleaned data in 2006 and 222 in 2007.

Week	Total deaths in raw data	Total deaths in cleaned data	Years in estimating sample	Total deaths in estimating sample
1	18584	14694	2007, 2008	2879
2	18300	14407	2007, 2008	2853
3	17940	14141	2007, 2008	2786
4	17575	13959	2007, 2008	2889
5	18035	14282	2007, 2008	2970
6	17543	13748	2007, 2008, 2009	4327
7	17583	13926	2007, 2008, 2009	4415
8	17525	13815	2007, 2008, 2009	4242
9	16327	12927	2007, 2008, 2009	4014
10	16690	13243	2007, 2008, 2009	4066
11	16347	12953	2007, 2008, 2009	3857
12	15866	12566	2007, 2008, 2009	3767
13	15605	12176	2007, 2008, 2009	3862
14	15620	12226	2007, 2008, 2009	3949
15	15630	12272	2007, 2008, 2009	3925
16	16041	12570	2007, 2008, 2009	3911
17	16345	12889	2007, 2008, 2009	3825
18	16334	12796	2007, 2008, 2009	3934
19	16407	12911	2007, 2008, 2009	3944
20	16713	13132	2007, 2008, 2009	4004
21	16949	13265	2007, 2008, 2009	4082
22	17067	13417	2007, 2008, 2009	4247
23	16650	13036	2007, 2008, 2009	4203
24	16379	12916	2007, 2008, 2009	4058
25	16765	13119	2007, 2008, 2009	4069
26	17047	13426	2007, 2008, 2009	3988
27	17308	13592	2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	12281
28	17754	13992	2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	12514
29	18392	14555	2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	13090
30	18406	14508	2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	14508
31	18520	14593	2006, 2007, 2008, 2009, 2010, 2011, 2012, 2014, 2015	13255
32	19210	14992	2006, 2007, 2008, 2009, 2010, 2011, 2012, 2014, 2015	13615
33	19418	15173	2006, 2007, 2008, 2009, 2010, 2011, 2012, 2014, 2015	13827
34	18910	14803	2006, 2007, 2008, 2009, 2010, 2011, 2012, 2014, 2015	13525
35	18507	14324	2006, 2007, 2008, 2009, 2010, 2011, 2014, 2015	11649
36	18744	14565	2006, 2007, 2008, 2009, 2010, 2011, 2014, 2015	11940
37	19310	15012	2006, 2007, 2008, 2009, 2010, 2011, 2014, 2015	12349
38	18843	14636	2006, 2007, 2008, 2009, 2010, 2011, 2014, 2015	11964
39	18601	14437	2006, 2007, 2008, 2009, 2010, 2011, 2014, 2015	11741
40	18240	14309	2006, 2007, 2010, 2011, 2014	7268
41	18037	13972	2006, 2007, 2010, 2011, 2014	7206
42	18676	14538	2006, 2007, 2010, 2011, 2014	7153
43	18131	14182	2006, 2007, 2010, 2011, 2014	7105
44	17959	14006	2006, 2007, 2011	4281
45	17432	13654	2006, 2007, 2011	4121
46	17012	13272	2006, 2007	2716
47	16691	13034	2006, 2007	2590
48	16858	12967	2006, 2007	2572
49	16878	13066	2006, 2007	2695
50	17068	13329	2006, 2007	2801
51	17391	13700	2006, 2007	2699
52	17830	14092	2006, 2007	2721
53	3103	2442	NA	NA

Table B.3: Years in estimating sample by week of year

Table shows a breakdown of the number of deaths in our raw data compared to the number of deaths in our estimating sample by week of year. Total deaths in cleaned data are lower than total deaths in raw data due to missing PIN codes in the addresses of some observations. Out of the total deaths in the cleaned data, the total deaths used in the estimating sample is determined by the availability of the AWS data for a given week in a given year.

C Regression Tables

C.1 Response functions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Age			Gender		Slum residence	
Daily Rainfall (mm)	All	<5	5-65	65+	Male	Female	Non-Slum	Slum
0	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-
50	0.0067	0.016	-0.00086	0.011	0.0056	0.008	0.0022	0.0082
	(0.0033)	(0.012)	(0.005)	(0.0042)	(0.0051)	(0.0049)	(0.005)	(0.0048)
100	0.013	0.03	0.004	0.018	0.0097	0.018	0.0058	0.017
	(0.0044)	(0.017)	(0.0073)	(0.0055)	(0.0065)	(0.0073)	(0.0061)	(0.0065)
150	0.022	0.046	0.015	0.024	0.014	0.032	0.011	0.029
	(0.0063)	(0.026)	(0.01)	(0.0081)	(0.0084)	(0.0096)	(0.009)	(0.0089)
250	0.05	0.096	0.054	0.041	0.034	0.073	0.023	0.075
	(0.013)	(0.044)	(0.02)	(0.015)	(0.016)	(0.016)	(0.016)	(0.019)
300	0.074	0.14	0.083	0.057	0.053	0.1	0.029	0.11
	(0.017)	(0.046)	(0.027)	(0.015)	(0.02)	(0.019)	(0.019)	(0.024)
N	19880		58158		39387		39288	
Sqr. corr.	0.88		0.83		0.8		0.85	

Table C.1: Percent change in 5-week mortality due to rainfall

Table reproduces results from the response functions depicted in main text figures. The first column, "All", shows the average response function over the whole sample, from Figure 5, Panel A. Columns 2 to 4 show the response functions from Figure 6, Panel A (estimated for 3 age groups in a stacked regression). Columns 5 and 6 show results from Figure 6, Panel B (estimated for both genders in a stacked regression). The final two columns show the results from Figure 7 Panel A (estimated for slum and non-slum residents in a stacked regression). Estimates are from models that use a third-order polynomial in daily rainfall for days in the contemporaneous week and 4 lagged weeks. Point estimates are shown for the percent change in 5-week mortality due to a single day with the total rainfall value indicated in each row, relative to a day with zero rainfall. Standard errors are in parentheses. Sqr. corr. refers to the squared correlation coefficient.

	(1)	(2)	(3)	(4)
		Linear tide interaction		
Hourly Rainfall (mm)	All	Tide Quantile: 5%	Tide Quantile: 50%	Tide Quantile: 95%
0	-	-	-	-
	-	-	-	-
5	7e-04	-0.0038	0.0029	0.009
	(0.00032)	(0.0042)	(0.0016)	(0.0037)
10	0.0015	-0.0063	0.0061	0.018
	(0.00049)	(0.0069)	(0.0025)	(0.0058)
15	0.0023	-0.0075	0.0098	0.025
	(0.00063)	(0.0099)	(0.0033)	(0.0077)
20	0.0032	-0.0071	0.014	0.033
	(0.00084)	(0.014)	(0.0043)	(0.01)
25	0.0043	-0.005	0.018	0.04
	(0.0011)	(0.018)	(0.0059)	(0.013)
30	0.0055	-0.0012	0.023	0.046
	(0.0016)	(0.022)	(0.0079)	(0.016)
N	19880		19880	
Sqr. Corr.	0.88		0.88	

Table C.2: Percent change in 5-week mortality due to hourly rainfall

Table reproduces the results from the hourly rainfall models presented in main text figures. All estimates use a third-order polynomial in hourly rainfall for hours in the contemporaneous week and 4 lagged weeks. Point estimates are shown for the percent change in 5-week mortality due to a single hour with the total rainfall value indicated in each row, relative to an hour with zero rainfall. Column 1 presents the average response function over the whole sample, from Figure 8. Columns 2-4 present response functions evaluated for the tide level at the quantile of the in-sample hourly tide distribution indicated in each column, reproducing the facets in Figure 9. Standard errors are in parentheses. Sqr. corr. refers to the squared correlation coefficient.

C.2 Cross-sectional relationship between elevation and slum status

To investigate whether slums are located in the low elevation areas within PIN codes, we compile a pixel level dataset recording each pixel's average elevation, and whether it is listed as a slum location by the BMC.[65]

In Table C.2, we present results from regressing elevation on a binary indicator for whether each pixel is classified as a slum.

In columns (1) and (2), we see that on average, slums are located in higher elevation areas of Mumbai. In column (2), we weight the regression using GHSL derived measures of each pixel's population.

[65]We obtained a shapefile used by Tandel et al. (2022) for the locations of slums in Mumbai through direct correspondence with the authors.

However, this relationship becomes small and statistically insignificant once we condition on PIN code fixed effects (columns 3 and 4). Therefore on average, slum location within a PIN code is not strongly correlated with elevation.

Table C.3: Elevation and slum location

Dependent Variable:	Elevation (m)			
Model:	(1)	(2)	(3)	(4)
Variables				
Slum Indicator	19.63***	17.40***	-4.133	0.7899
	(2.602)	(2.116)	(5.199)	(0.9322)
Fixed-effects				
Pincode			Yes	Yes
Regression weights	None	Population	None	Population
Fit statistics				
Observations	51,706	33,950	51,706	33,950

Clustered (pincode) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

D Robustness checks

Throughout, we colour our preferred specification, which is used in the main text, in blue, and robustness checks in grey.

D.1 Weekly vs. daily resolution

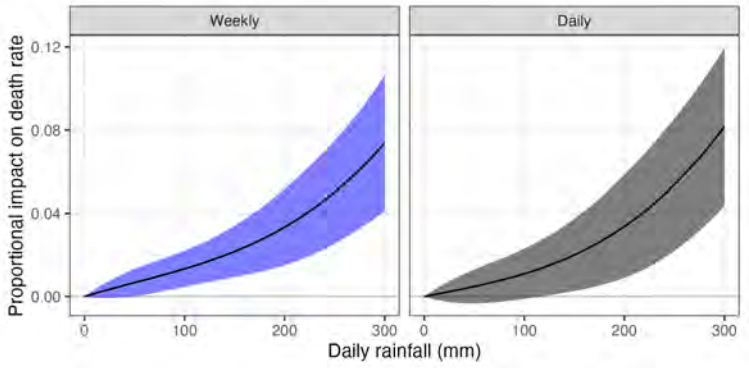

Figure D.1: Percent change in 5-week mortality due to daily rainfall: estimated at the daily vs weekly level.

Mortality-rainfall response functions are estimated for the full sample. The left hand side shows our main specification, which is a weekly level regression with 4 weekly lags. The right hand panel shows results from a model that is run at the daily level, and includes a 3rd order polynomial in contemporaneous rainfall, and 35 daily lags. Points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals.

D.2 Polynomial choice

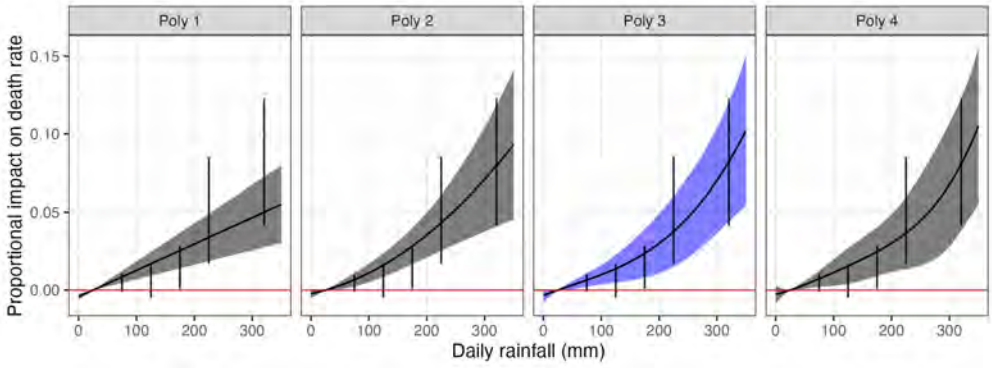

Figure D.2: Percent change in 5-week mortality due to daily rainfall: various polynomial and binned functional forms.

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that use polynomials in daily rainfall for days in the contemporaneous week and 4 lagged weeks. Black error bars represent the results from a regression that uses bins of the daily rainfall distribution as the functional form of rainfall (Deschênes and Greenstone, 2011). Bins are defined as $\{[0, 50), [50, 100), [100, 150), [150, 200), [200, 250), [250, \infty)\}$ mm of daily rainfall. Overlaid curves show results from models with polynomial degree indicated in the facet titles. Points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas and error bars indicate 95% confidence intervals.

D.3 Level of temporal aggregation before taking non-linearity

Here we explore the implications of taking non-linear transformations of rainfall at different levels of temporal aggregation— hourly, daily, or weekly.

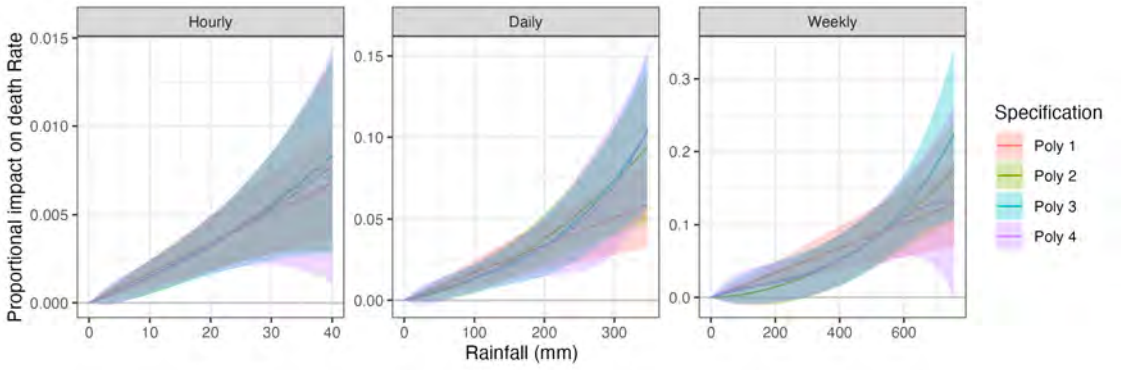

Figure D.3: Percent change in 5-week mortality due to daily rainfall: non-linearity taken at hourly vs daily vs weekly level.

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses polynomials in hourly, daily, or weekly rainfall for periods in the contemporaneous week and 4 lagged weeks. Overlaid curves show results from models with polynomial degree indicated in the plot legend. Points along each curve represent the percent change in 5-week mortality due to a single period (hour/day/week) with the total rainfall value shown on the x-axis, relative to a period with zero rainfall. Shaded areas indicate 95% confidence intervals.

We focus on the daily model, since we believe this captures non-linearities at an economically meaningful level. The hourly model appears roughly linear, and thus non-linearities at a very high-frequency do not seem to be driving the non-linearity at the daily level. We find the weekly model is noisier and less stable across polynomial orders.

D.4 Spatial and temporal spillovers

Urban flooding impacts mortality through complex mechanisms. These include direct injuries, vector- and water-borne diseases and economic disruption. Our dataset allows us to run our analysis at a high temporal and spatial resolution. However, the high level of resolution brings with it the likelihood that our estimates are biased because they fail to account for temporal and spatial spillover effects. In this subsection, we investigate these possibilities, and conclude that our main specification likely understates the true causal effect due to long lasting temporal spillovers. We do not find meaningful evidence of spatial spillovers.

D.4.1 Temporal spillovers

Our main specification of the mortality-rainfall response (Equation (2)) includes rainfall variables for the contemporaneous week and 4 lagged weeks, thereby allowing rainfall in

a given week to affect mortality in the contemporaneous week and 4 subsequent weeks. Figure 5 Panel A in the main text depicts the effect of a day with a given amount of rainfall on mortality over this entire 5 week period. In Figure 5 Panel B, we present the breakdown of the effect for each of the 5 weeks and see that it is roughly similar in each week.[66] This suggests that while there are immediate mortality effects of extreme rainfall in the contemporaneous week, these effects also persist into subsequent weeks, consistent with flood-related diseases taking time to develop and spread. The increased mortality in each of the subsequent weeks reveal that rainfall shocks do not simply advance the timing of deaths without altering the total number of deaths.

To further investigate the temporal structure of the causal impacts of extreme rainfall, we utilise our data at the daily level to explore the lag structure of our estimated effects.

Figure D.4 shows the results from estimating regressions of daily mortality on daily rainfall, varying the number of lags of daily rainfall variables included.[67] The left hand-side panel shows the summed lag response function from models with 5, 35, 75 and 95 daily lags. The right hand side panel shows the point estimate and standard error of the response function for 100, 200, and 300mm of daily rainfall for models with between 0 and 100 daily lags, where the response is obtained by summing over the lags.

[66] The height of the curve in Figure 5 Panel A is equal to the average of the height of the 5 curves in Figure 5 Panel B. The averaging is necessary to aggregate percent increases in each week's mortality into a percent increase in mortality over a 5-week period.

[67] These regressions include PIN code $\times$ year fixed effects and date (e.g., July 14, 2013) fixed effects.

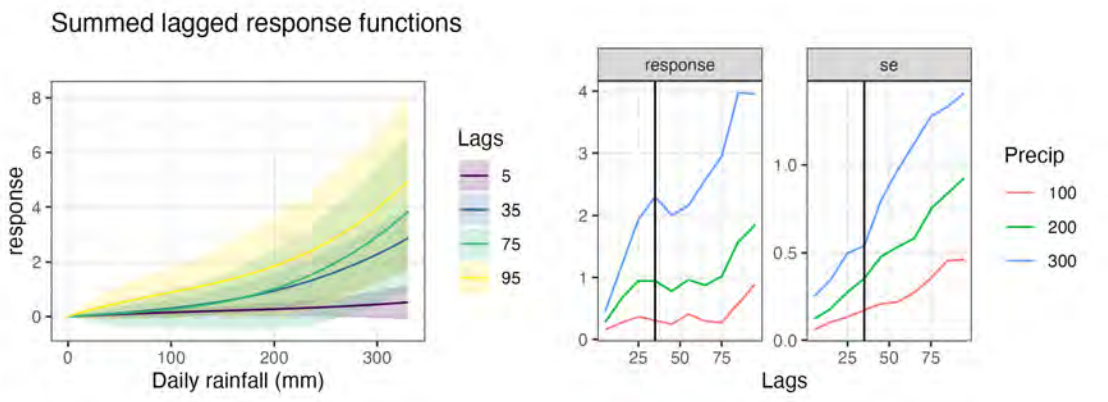

Figure D.4: Percent change in mortality due to daily rainfall: daily regressions with increasing numbers of lags.

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third order polynomial in daily rainfall for the contemporaneous daily and various numbers of lags. Points along each curve represent the percent change in mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. Unlike in other specifications, we here do not normalise to represent impacts over a five week period. The left hand panel shows the sum over contemporaneous and lagged responses at each level of daily rainfall indicated on the x-axis. We find that when including more lags, the mortality response to a given level of rainfall is noisier, but larger in magnitude. The right hand panel shows the point estimate and standard error of the cumulative response function (summed over contemporaneous and lagged responses) at 100 mm, 200 mm, and 300 mm rainfall, for specifications with varying numbers of lags (indicated on the x-axis). Our main specification, uses 35 days worth of lags (i.e., 5 weeks), which is indicated by the black vertical line.

If extreme rainfall causes a deterioration in the health of individuals who are already weak and would have died even in the absence of the rainfall shock, then we should expect the sum of the lagged response functions to eventually start tending down towards zero, as the only effect of the extreme rainfall would be to alter the timing of mortality but not the number of deaths in the longer run. However, we do not see any evidence for this narrative in our data. Instead the mortality response is remarkably persistent over time, which is consistent with a diseases developing and spreading over longer time horizons. Figure D.4 suggests that a specification that accounts for roughly 35 days of lagged rainfall is a lower bound on the true magnitude of the response function, as elevated mortality appears to persist considerably longer than that. We note here that a regression of daily mortality that includes 35 days of lagged rainfall variables is analogous to our main specification, where we regress weekly mortality on four lagged weeks' in addition to the contemporaneous week's rainfall variables.

Our results would likely have a larger magnitude if we included more lags. However, including more lags would reduce the number of observations available to us in our analy-

sis, resulting in less statistical power, especially when modelling heterogeneous responses along various dimensions (i.e., age, slum residence, elevation). We therefore choose a parsimonious model with fewer lags, but note that this model likely understates the true causal effect.

D.4.2 Spatial spillovers

Another potential concern with our estimates is that we fail to account for the fact that the impacts of rainfall likely extend beyond the PIN code in which it occurs. This is both due to the drainage patterns in the city, and the reality that people spend a lot of time outside of the PIN code that contains their residential address.

While this is a valid concern, a few points are worth noting. First, these spillovers likely bias our estimates downwards. This is because additional effects of rainfall shocks in PIN codes other than those that they directly affect are likely in the same direction as our estimates. Second, even in the presence of these spillovers, we can still interpret our estimates as being the effect of extreme rainfall, net of spillovers and avoidance behaviours.

Nevertheless, we implement two strategies to investigate the importance of spatial spillovers in our analysis. First, we estimate the response function at various levels of spatial aggregation. Second, we implement a specification in which we separately control for rainfall in neighbouring PIN codes.

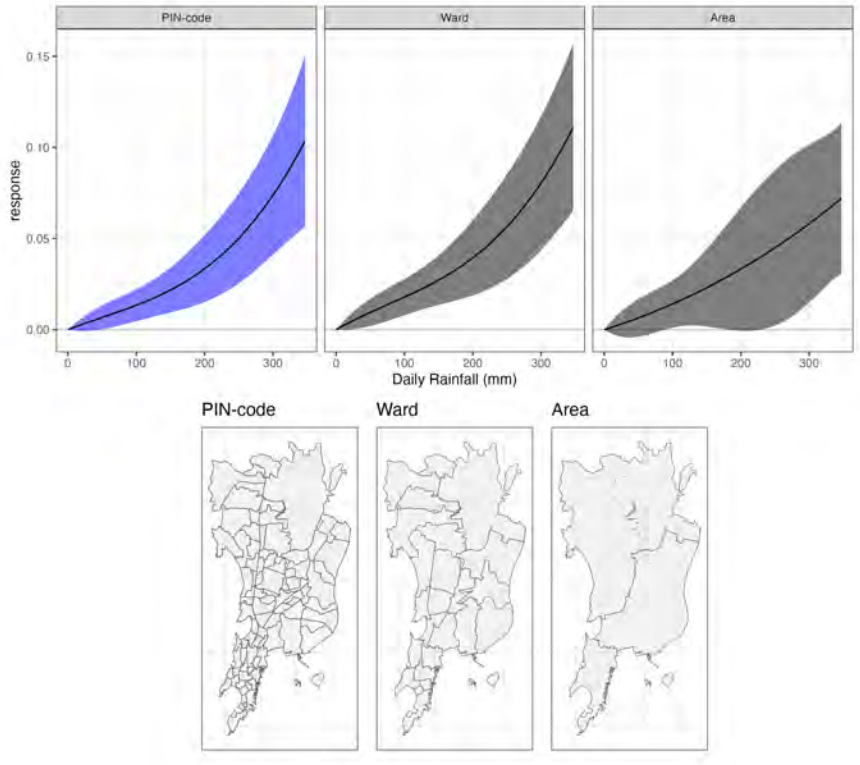

Figure D.5: Percent change in 5-week mortality due to daily rainfall: estimated at various levels of spatial aggregation.

Mortality-rainfall response functions are estimated for the full sample. All panels show estimates from a weekly level regression that uses a third order polynomial in daily rainfall. Points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. The left hand panel reproduces our main specification, from Panel (A) of Figure 5. The middle panel aggregates PIN codes to the ward with which they share the most overlap, and clusters at the ward and date level. The right hand panel shows the response function from a model estimated at the "area" level (i.e., approximate watersheds), where areas are as denoted in the map below. Since there are only three spatial units in this regression, we cannot cluster at the spatial unit level, and only cluster by date. The maps below show the associated spatial units for each plot.

From Figure D.5, we conclude that spillovers at the ward or area level do not greatly affect our results, since the response function is qualitatively similar when estimated at each of these levels of aggregation. We define an "area" of Mumbai, by manual inspection of maps of Mumbai's watersheds. This grouping is chosen to try and capture sets of PIN codes that are likely to experience spillovers due to runoff from each-other's rainfall shocks.

We are not able to run this analysis at the city level, as we would not be able to control for time fixed effects which are crucial for our identification argument.

We next implement a version of Equation (1) that controls for the mean rainfall level in neighbouring PIN codes. Note, all specifications include time fixed effects, which control

for the average level of rainfall across all PIN codes in each period in the sample.

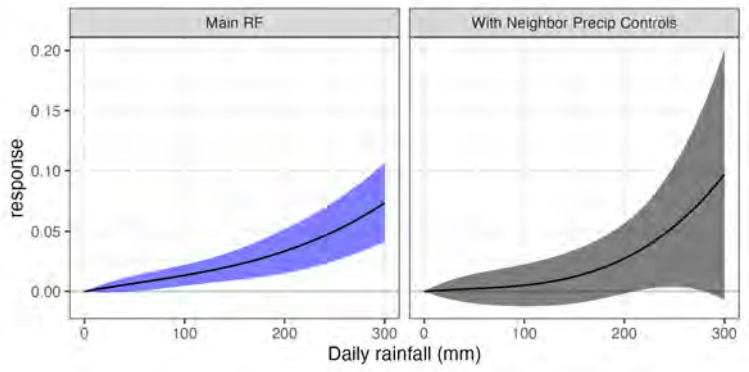

Figure D.6: Percent change in 5-week mortality due to daily rainfall: controlling for neighbouring rainfall

Mortality-rainfall response functions are estimated for the full sample. The figure shows the estimated rainfall response function from two models, both of which use a third-order polynomial in the own-PIN code daily rainfall in the contemporaneous week and four lagged weeks. The model on the left panel reproduces our main specification, from Panel (A) of Figure 5. The right panel shows a model that additionally contains a third-order polynomial in the neighbouring PIN code's daily rainfall in the contemporaneous week and four lagged weeks. Points along each curve represent the percent change in 5-week mortality due to a single day with the total own-PIN code rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals.

Controlling for neighbouring precipitation increases the uncertainty associated with the estimated effects of own-PIN code rainfall. However, the magnitude and qualitative shape of our estimated mortality response to own-PIN code rainfall remains robust (Figure D.6). Figure D.7 compares the mortality effects of own-PIN code and neighbouring PIN code rainfall. The effects of neighbouring PIN code rainfall are very imprecisely estimated.

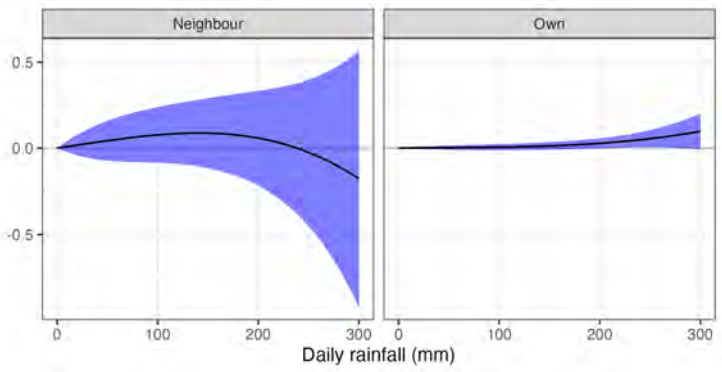

Figure D.7: Percent change in 5-week mortality due to daily rainfall: own vs. neighbouring PIN code rainfall

Mortality-rainfall response functions are estimated for the full sample, using a model that contains a third-order polynomial in own and neighbouring PIN code's daily rainfall in the contemporaneous week and four lagged weeks. Points along each curve represent the percent change in 5-week mortality due to a single day with the total own-PIN code rainfall value (right panel) or neighbouring PIN code rainfall value (left panel) shown on the x-axis, relative to a day with zero rainfall. The right panel reproduces the mortality response to own-PIN code rainfall that is shown in Figure D.6, right panel. Shaded areas indicate 95% confidence intervals.

Overall, we conclude that while spatial spillover effects are important to consider when interpreting our results, our results do not change substantially when controlling for spillovers. We therefore focus on a simple model that considers PIN codes as independently treated units.

D.5 Estimating sample

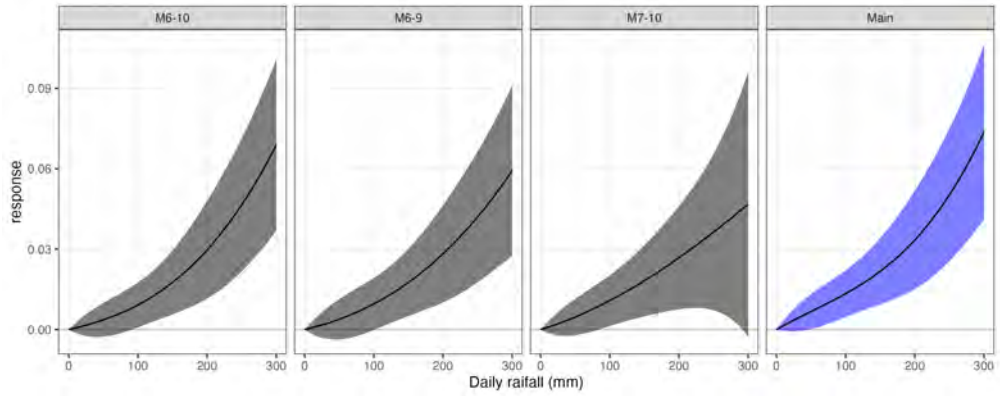

Figure D.8: Percent change in 5-week mortality due to daily rainfall: month of year sub-samples.

Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals.

The rightmost panel reproduces our main specification, from Panel (A) of Figure 5. The other panels re-estimate the response function on sub-samples of the months of the year indicated in the facet titles. For example, the leftmost panel shows the response function estimated on a sub-sample of our regression data where data from all months outside of June (month 6) to October (month 10) are dropped.

D.6 Estimation strategy

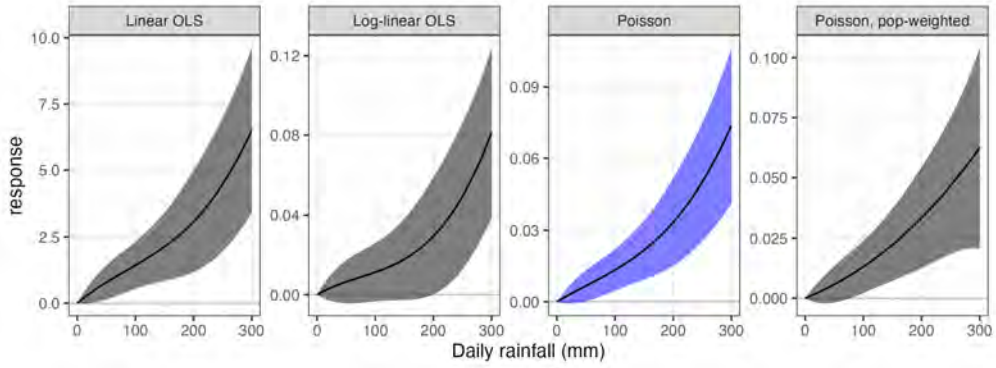

Figure D.9: Change in 5-week mortality due to daily rainfall: various estimation strategies. Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. Facet titles indicate estimation strategy. Linear OLS indicates a model run in levels rather than logs, with the mortality count as the dependent variable. Log-linear OLS requires dropping observations with zero deaths. Poisson, the third panel, is our main specification, reproducing Panel (A) of Figure 5. The final panel displays the estimates from a Poisson model with PIN code level regression weights, where the weights are proportional to the 2005 PIN code level population (calculated using GHSL satellite data).

D.7 Temperature controls

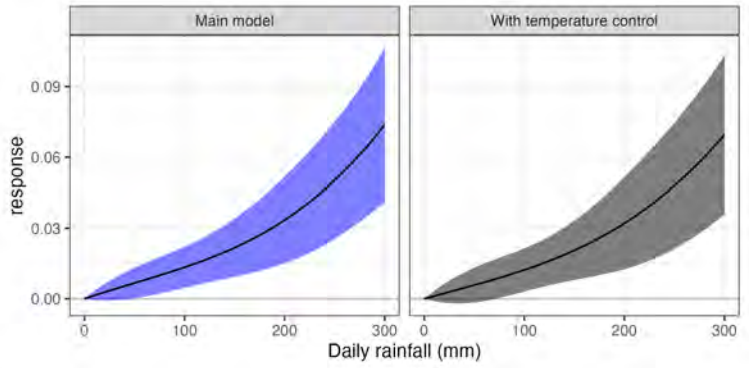

Figure D.10: Percent change in 5-week mortality due to daily rainfall: with and without temperature controls.

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. The left hand panel reproduces our main specification, from Panel (A) of Figure 5. The right hand panel displays estimates of the mortality-rainfall response from a model that additionally includes temperature controls, where the temperature response function is modelled as 3rd order polynomial in the daily average temperature in the contemporaneous week and four lags, analogously to the rainfall response function.

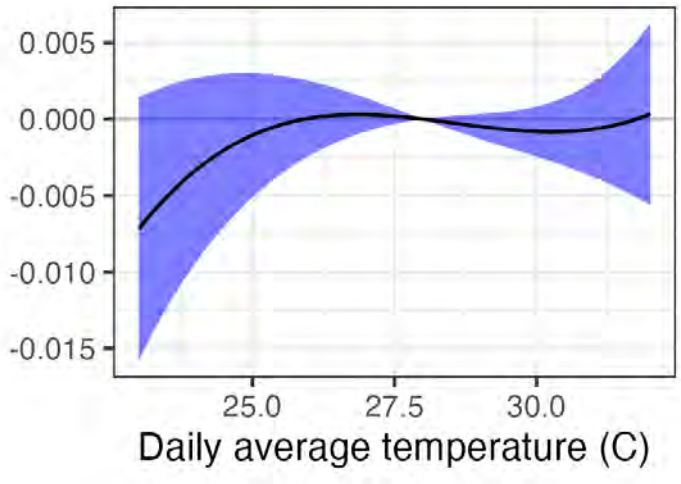

Figure D.11: Percent change in 5-week mortality due to daily temperature.

Mortality-temperature response function is estimated for the full sample. Estimates are from a model that uses a a third-order polynomial in daily average temperature in the contemporaneous week and four lags, and a third-order polynomial in daily rainfall in the contemporaneous week and four lags (same model as shown in the right panel of Figure D.10). Points along each curve represent the change in 5-week mortality due to a single day with the daily average temperature value shown on the x-axis, relative to a day with daily average temperature of 28°C. Shaded areas indicate 95% confidence intervals.

D.8 Location assignment

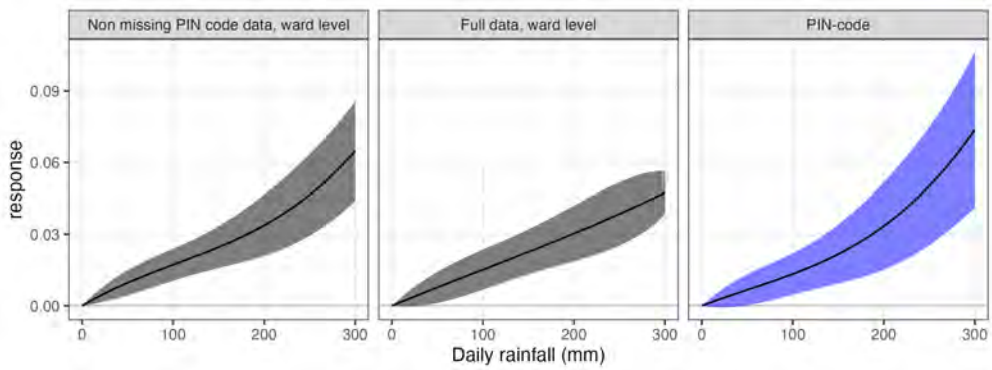

Figure D.12: Percent change in 5-week mortality due to daily rainfall: assigning location using different methods

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. The rightmost panel reproduces our our main specification, from Panel (A) of Figure 5, in which we match rainfall to mortality records by using the Mumbai PIN code information present in about 80% of the addresses. The leftmost panel uses the same data as our main specification (i.e. we drop the same 20% of data for which we have no PIN code match), but uses the municipal ward assigned to the mortality record as the location identifier. The middle panel also uses the ward identifier, but does not drop the data that does not have a Mumbai PIN code.

D.9 Choice of fixed effect

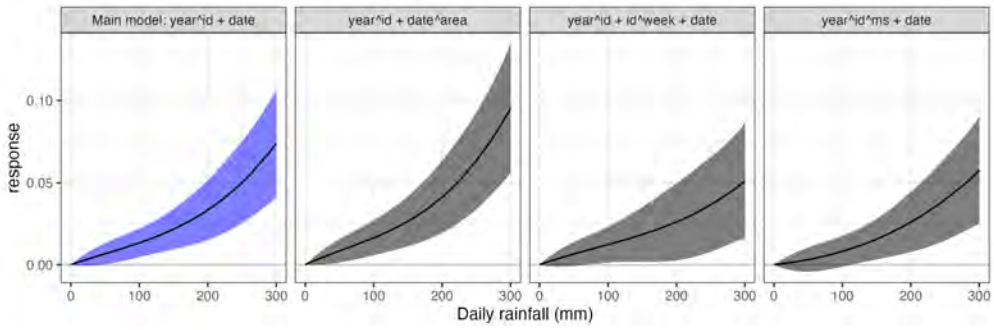

Figure D.13: Percent change in 5-week mortality due to daily rainfall: various fixed effects
Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. The leftmost panel reproduces our main specification, from Panel (A) of Figure 5, in which we use PIN code $\times$ year and week $\times$ year (i.e., date) fixed effects. The second-from-left panel further interacts the week $\times$ year fixed effects with an indicator representing three areas of Mumbai, which are the same as those used in Figure D.5. The second-from-right panel uses PIN code $\times$ year, PIN code $\times$ week-of-year, and week $\times$ year fixed effects. Finally, the rightmost panel interacts the PIN code $\times$ year fixed effects with an indicator for whether the month-of-year is between June and September (inclusive), and also uses week $\times$ year fixed effects.

D.10 Alternative AWS cleaning steps

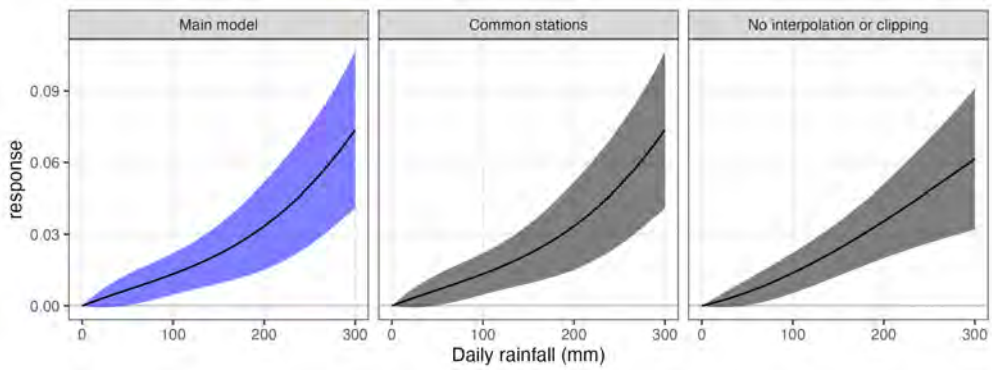

Figure D.14: Percent change in 5-week mortality due to daily rainfall: various AWS data cleaning procedures

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. The left hand panel reproduces our our main specification, from Panel (A) of Figure 5, in which we clean the AWS data as outlined in Appendix Section B.3. The middle panel restricts the set of stations used to calculate our PIN code level averages to the 13 stations that are available across all years in our sample. Finally, the right hand panel doesn't spatially or temporally interpolate any station data, or clip extreme implausible values, before calculating the PIN-code level spatial averages of the available stations.

D.11 Randomisation check

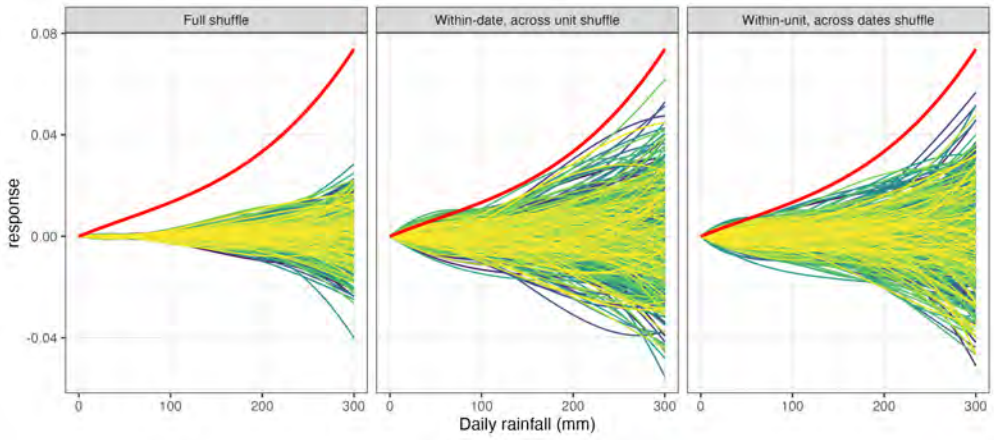

Figure D.15: Percent change in 5-week mortality due to daily rainfall: randomisation check
Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Shaded areas indicate 95% confidence intervals. The red line presents our main specification, from Panel (A) of Figure 5. Other lines show estimated response functions from data where the weather data is shuffled across space and or time, as a placebo check to rule out that spurious correlations are not driving our results (Hsiang and Jina, 2014). Each plot shows the results from 500 models. The leftmost panel shuffles rainfall values randomly across both dates and PIN codes. The middle panel shuffles full weather histories across PIN codes, but preserves the time series nature of the treatment. The rightmost panel shuffles dates across time within PIN codes. In each plot, we see our red estimated response function is on the outer-envelope of the randomised estimates, indicating it is very unlikely that our estimated response function occurred purely by chance or due to spurious correlations.

E Additional heterogeneity

E.1 Age categories

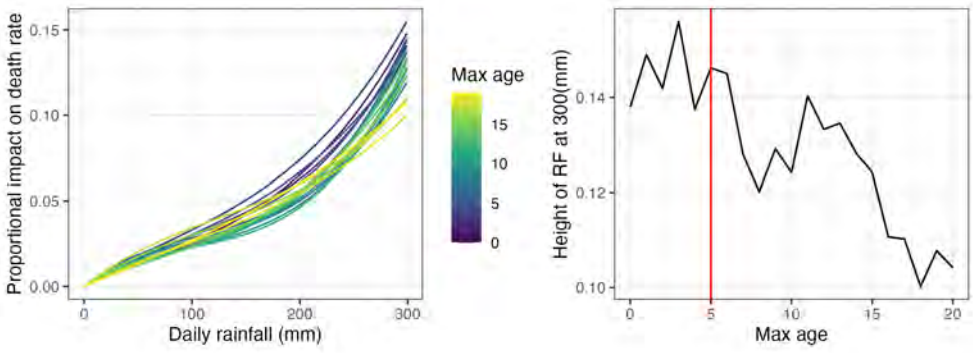

Figure E.1: Percent change in 5-week mortality due to daily rainfall: robustness under different categorisations of the minimum age group

Mortality-rainfall response functions are estimated for the full sample, using Equation (3) that allows for heterogeneity across age categories. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curves in the left hand side plot represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Each line in the left hand plot shows the estimated response function for the youngest age category, where the youngest age category is $[0, X]$, $X \in [0, 20]$. Qualitatively, we can see that the strongest relationship is found when we restrict the category to only include younger people (the bluer lines). The right hand plot shows the height of the response function at 300mm as a function of the maximum age in the youngest age category. The red line at 5 years denotes our main specification, where the youngest age category consists of persons under 5 years of age.

E.2 Age × gender

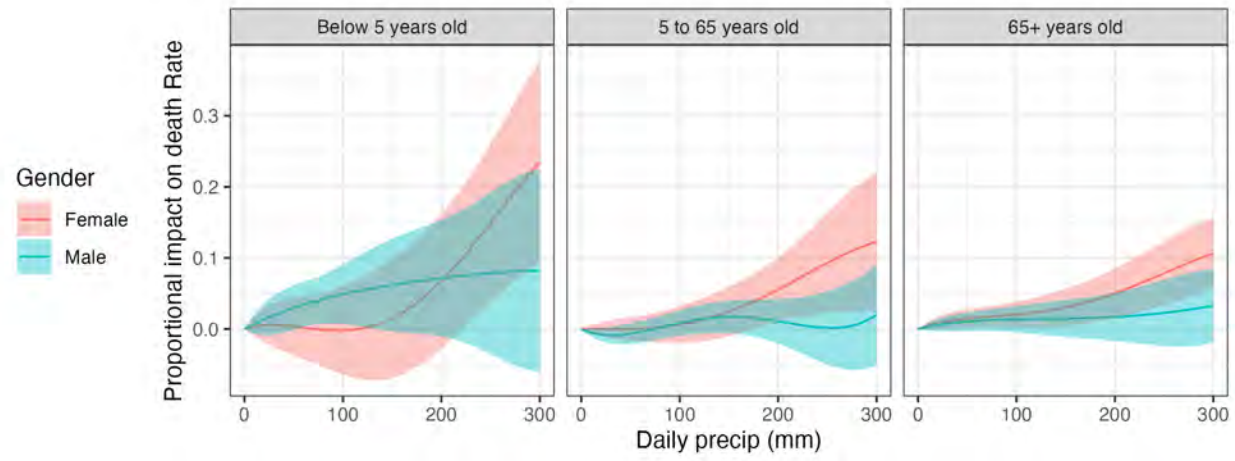

Figure E.2: Percent change in 5-week mortality due to daily rainfall: heterogeneity by age and gender.

Mortality-rainfall response functions are estimated for the full sample, using Equation (3) that allows for heterogeneity across age categories and gender. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curves in the left hand side plot represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall.

E.3 Slum residence × age category

Category	Number of deaths	Proportion of total deaths
Below 5, non-slum	7693	2.1
Below 5, slum	17160	4.8
5 to 65, non-slum	64500	17.9
5 to 65, slum	101210	28.1
65+, non-slum	92543	25.7
65+, slum	77131	21.4

Table E.1: Deaths by age group and slum/non-slum residence.

Table shows the number and proportion of deaths in our estimating sample, for each of the slum-age categories.

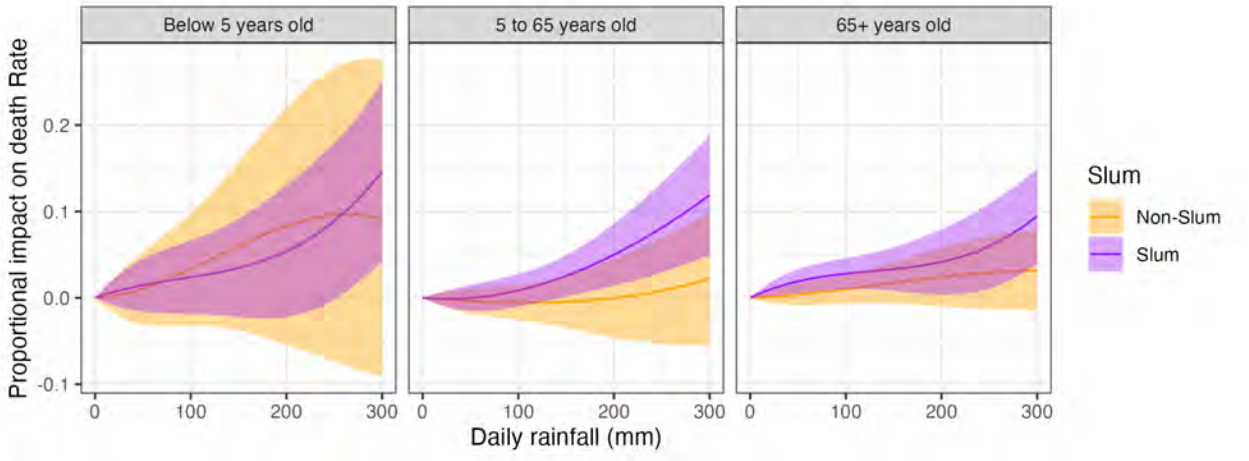

Figure E.3: Percent change in 5-week mortality due to daily rainfall: heterogeneity by age and slum residence.

Mortality-rainfall response functions are estimated for the full sample, using Equation (3) that allows for heterogeneity across age categories and slum vs non-slum residents. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curves in the left hand side plot represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall.

	Non-slum residence			Slum residence		
Total	<5 years	5-65 years	≥ 65 years	<5 years	5-65 years	≥ 65 years
0.069	0.13	-0.042	0.052	0.17	0.023	0.21
(-0.011, 0.17)	(-0.25, 0.85)	(-0.2, 0.15)	(-0.08, 0.21)	(-0.13, 0.68)	(-0.14, 0.22)	(0.053, 0.39)

Table E.2: Proportion of monsoon season deaths caused by rainfall shocks, by age × residence.

Table shows the proportion of total deaths and age- and residence-wise deaths during June to September that we attribute to rainfall shocks. Estimates are derived from a model with heterogeneous mortality-rainfall responses by age × slum residence (Figure E.3). Point estimates are calculated based on Equation (7); 95% confidence intervals, calculated with a parametric bootstrap described in Appendix F.1, are in parentheses.

E.4 Elevation

Given the importance of drainage patterns and topology in determining urban flooding, we expect that lower elevation PIN code are more at risk of damages from extreme rainfall patterns. To test this, we split the PIN codes in our sample into two groups according to either their mean elevation level, mean height above nearest drainage (HAND)[68], or percent of land area below 5 metres elevation.

[68] "Height Above Nearest Drainage" (HAND) is calculated by Tripathy et al. (2020) and obtained through direct correspondence.

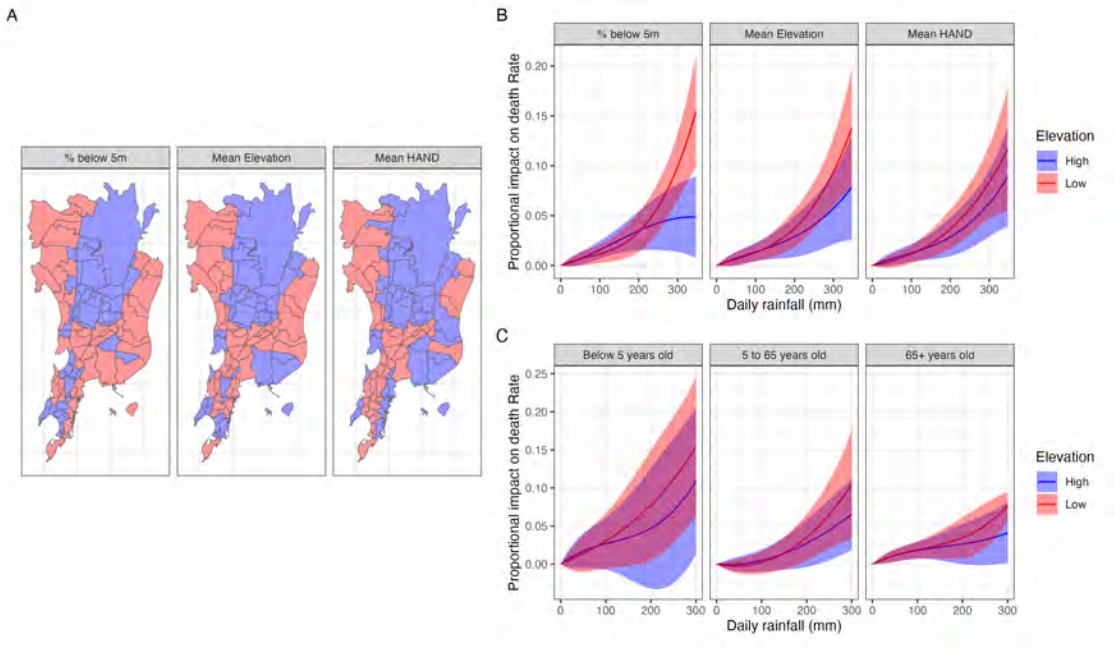

Figure E.4: Percent change in 5-week mortality due to daily rainfall: various definitions of elevation heterogeneity.

Mortality-rainfall response functions are estimated for the full sample, using a modified version of Equation (1) that allows for heterogeneity across high and low elevation PIN codes. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curves in the left hand side plot represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Panel (A) shows the spatial distribution of high and low elevation PIN codes, according to three different definitions of PIN-code level elevation. For each definition, we take above and below median values of a particular PIN code level statistic to define whether the PIN-code is high or low elevation. '% below 5m' refers to the proportion of pixels in each PIN code that are below 5m above sea-level. 'Mean Elevation' is our main measure of elevation, and simply denotes the mean elevation across pixels within each PIN code. 'Mean HAND' denotes the mean height above nearest drainage for each pixel in each PIN-code. Panel (B) shows response function heterogeneity by elevation according to each of these measures. Finally, Panel (C) shows elevation by age heterogeneity, where each age category is allowed to have a separate response in high and low elevation PIN codes as defined by mean elevation.

E.5 Tide

E.5.1 450cm Tide threshold model

Let $\tau_{h,d,w,y}$ denote the hourly tide level on hour h of day d in week w and year y, and let $\bar{\tau}$ denote a threshold tide level. We allow for a separate response to hourly rainfall occurring when the hourly tide is below vs. above the threshold tide level (i.e., "low" vs. "high"):

$$m_{i,w,y} = g_{low}(\boldsymbol{R}^{low}_{i,w,y}) + g_{high}(\boldsymbol{R}^{high}_{i,w,y}) + \alpha_{i,y} + \lambda_{w,y}, \tag{E.1}$$

To construct the functions g_{low} and g_{high}, we account for nonlinear transformations and lags of rainfall in a similar manner as Equation (5). We construct non-linear transformations of rainfall at the hourly level, before aggregating to the weekly level across all low-tide hours (in the case of g_{low}) or high-tide hours (in the case of g_{high}):

$$g_{low}(\boldsymbol{R}^{low}_{i,w,y}) = \sum_{k=1}^{3}\sum_{l=0}^{4}\beta^{low}_{l,k}\sum_{d\in w-l}\sum_{h\in d}\left(R^{k}_{i,h,d,w-l,y}\cdot \mathbf{1}_{\tau_{h,d,w,y}<\bar{\tau}}\right),$$

$$g_{high}(\boldsymbol{R}^{high}_{i,w,y}) = \sum_{k=1}^{3}\sum_{l=0}^{4}\beta^{high}_{l,k}\sum_{d\in w-l}\sum_{h\in d}\left(R^{k}_{i,h,d,w-l,y}\cdot \mathbf{1}_{\tau_{h,d,w,y}\geq\bar{\tau}}\right),$$

The indicator variables $\mathbf{1}_{\tau_{h,d,w,y}<\bar{\tau}}$ and $\mathbf{1}_{\tau_{h,d,w,y}\geq\bar{\tau}}$ respectively indicate whether the maximum tide level in hour h was below or above the threshold level $\bar{\tau}$.

We estimate Equation (E.1) where $\bar{\tau}$ is the threshold value of 450 cm, for which rainfall is expected to have larger impacts. Figure E.5 shows the mortality response to hourly rainfall occurring during a tide height of below 450 cm in the right panel (i.e., $g_{low}()$ in Equation (E.1)), and during a tide height of above 450 cm in the left panel (i.e., $g_{high}()$ in Equation (E.1)). The average response function, presented in Figure 8, is included for reference in green in both panels.

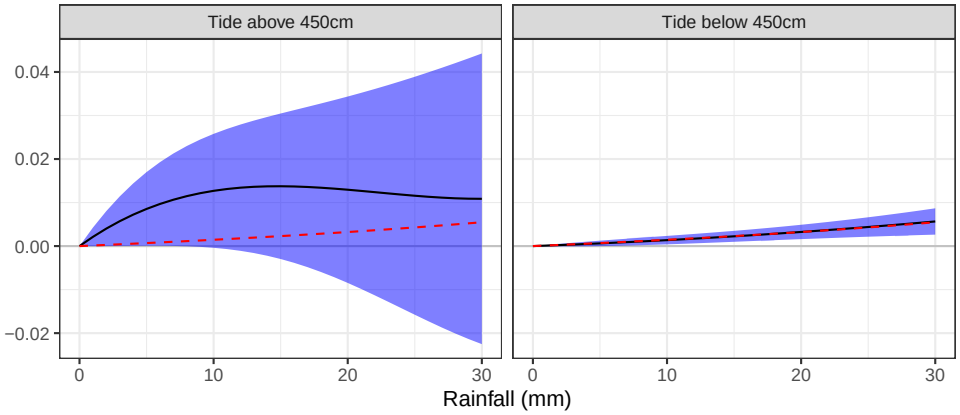

Figure E.5: Percent change in 5-week mortality due to hourly rainfall occurring during low and high tide periods.

Separate mortality-rainfall response functions are estimated for total hourly rainfall occurring during periods of the day when the tide height is above 450 cm (left panel, blue curve) and below 450 cm (right panel, blue curve). Estimates are from a model that uses a third-order polynomial in hourly rainfall during below/above 450 cm tide heights, for days in the contemporaneous week and 4 lagged weeks (Equation (E.1)). Points along each blue curve represent the percent change in 5-week mortality due to a single hour with the total rainfall value during above 450 cm tide (left panel) or below 450 cm tide (right panel) shown on the x-axis, relative to an hour with zero rainfall during above 450 cm tide (left panel) or below 450 cm tide (right panel). Shaded areas indicate 95% confidence intervals. In each panel, the green curve reproduces the average effect of hourly rainfall on mortality that is shown in Figure 8.

We find qualitatively large, but noisy evidence of heterogeneity by high and low tide in this specification. Even a small amount of hourly rainfall (e.g., 5 mm) that happens during a tide above 450 cm is far more damaging than heavy rainfall that happens when the tide is below 450 cm. It should be noted that tides above 450 cm are relatively rare, occurring over only 1.6% of the total hours in a year, on average in our sample. However, the extreme mortality effects of rainfall that we find during these periods is consistent with reports that suggest low levels of rainfall during a high-tide can quickly overwhelm Mumbai's drainage system.

Table E.3 shows the results from the attribution exercise described in Section 5 for this model. Relative to the linear interaction model presented in the main text, this model leads us to attribute a substantially higher proportion of deaths to rainfall during high tide.

Model	Total	Low Tide Rainfall	High Tide Rainfall
Hourly rainfall	0.092 (0.024, 0.17)		
Hourly rainfall by extreme tide	0.1 (0.032, 0.18)	0.081 (0.017, 0.16)	0.019 (-0.00021, 0.04)
Proportion of rainfall	1	0.98	0.02
Proportion of attributed deaths	1	0.81	0.19

Table E.3: Proportion of monsoon season deaths caused by rainfall shocks: hourly rainfall models.

The first row shows the proportion of total deaths during June to September that we attribute to rainfall shocks, derived from a model with hourly rainfall (Equation (4)). The second row shows the proportion of total deaths during June to September that we attribute to rainfall shocks during hours with high tide ($\geq$ 450 cm) and low tide ($<$ 450 cm) and overall, derived from the model specified in Equation (E.1). All point estimates are calculated based on Equation (7), and 95% confidence intervals (in parentheses) are calculated with a parametric bootstrap described in Appendix F.1. The third row summarises the proportion of rainfall in the sample that occurs during hours in which the average tide value is low ($<$ 450 cm) or high ($\geq$ 450 cm). The fourth row states the proportion of attributed deaths in the second row that are caused by rainfall shocks during low tide or high tide.

E.5.2 Daily tide model

Our subdaily tide interaction models presented in main text Figure 9 and Figure E.5 show rainfall occurring during a high tide within a day can have an especially large impact on mortality.

It is also possible that flooding can occur when heavy rain falls during periods of high daily average tide height.[69] To investigate this, we linearly interact the daily total rainfall with a measure of the average daily tide value. Let $T_{d,w,y}$ denote the daily average tide level on day d, week w and year y, after normalising it to be mean zero and have standard deviation 1. We augment our basic response function (Equation (2)) to allow each variable to have different impacts at different levels of tide as follows:

[69] For example, the BMC notes that apart from high maximum tides, neap tides may also cause flooding in Mumbai when coupled with high rainfall. During a neap tide there is a narrow range between high and low tide heights, meaning drainage is more difficult. A neap tide occurs when the sun and moon are at right angles to each other, which weakens their gravitational pull on the Earth's oceans, resulting in a narrow tidal range. Neap tides occur during the first and third quarter moon. Figure 4 depicts tidal patterns in August 2008, illustrating fluctuations in tidal heights and amplitudes.

$$g^{\text{tide}}(\boldsymbol{R}_{i,w,y}) = \sum_{k=1}^{3}\sum_{l=0}^{4}\beta_{l,k}\sum_{d\in w-l}\left(R^{k}_{i,d,w-l,y}\right) + \sum_{k=1}^{3}\sum_{l=0}^{4}\gamma_{l,k}\sum_{d\in w-l}\left(T_{d,w-l,y}\cdot R^{k}_{i,d,w-l,y}\right), \quad \text{(E.2)}$$

where the interactions terms between rainfall variables and tide (with γ coefficients) are added to the uninteracted rainfall variables from Equation (2).

Figure E.6 presents the results of estimating this model. We find that when the average tide level of a day is higher, there is a stronger impact of rainfall on mortality. We reject the joint null hypothesis that all γ terms are zero at the 1% level.

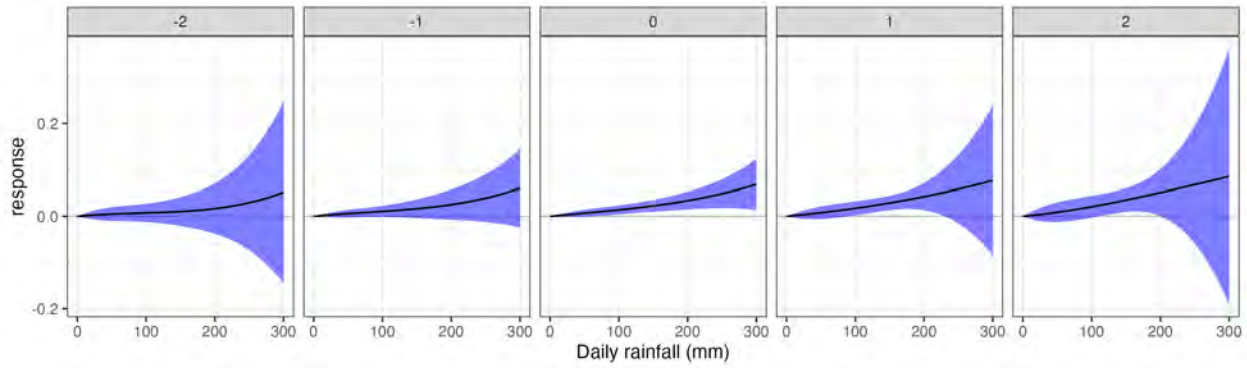

Figure E.6: Percent change in 5-week mortality due to daily rainfall: linear daily tide interaction.

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall and a linear tide-rainfall interaction, in the contemporaneous week and four lags. Points along each curves in the left hand side plot represent the change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Each facet shows the estimated response function evaluated at different levels of the normalised daily tide value. When daily tide is very low (e.g., two standard deviations below the mean, leftmost panel), the mortality impacts of rainfall are muted, while when daily tide is very high (e.g., two standard deviations above the mean, right hand panel) rainfall causes much larger mortality impacts.

F The mortality burden of rainfall shocks: Uncertainty and robustness

F.1 Uncertainty in estimates of mortality due to rainfall shocks

The procedure for estimating uncertainty for our attribution analysis is as follows:

1. Estimate the model, save the coefficients and variance-covariance matrix. Save the baseline values, $exp(FE)$. Calculate the central estimate for the $\hat{r}$.
2. Take 1000 draws from the joint normal sampling distribution of the coefficient estimates.
3. Calculate $\hat{r}^k_{i,w,y}$ for each bootstrap draw k for each observation in our sample.
4. Sum these estimates across weeks in each month in our sample, for each bootstrap draw k.
5. Take quantiles of the distribution of $\hat{r}^k_{i,m}$ to get estimates for the upper and lower confidence interval.
6. Divide upper and lower confidence interval values by the total deaths observed in each months in order to express our values as proportions.

F.2 Mortality due to rainfall shocks under alternative fixed effects

In Table F.1 we present estimates of the proportion of monsoon season deaths caused by rainfall in our main model (first row), and in three alternative robustness models with more saturated fixed effects (second, third, and fourth rows).

Fixed effect	Estimate
year^PIN code + date	0.089 (0.013, 0.17)
year^PIN code + date^area	0.1 (0.024, 0.19)
year^PIN code^Monsoon Season + date	0.046 (-0.044, 0.14)
year^PIN code + PIN code^week + date	0.085 (0.0019, 0.18)

Table F.1: Proportion of monsoon season deaths caused by rainfall shocks in alternative empirical specifications.

Table shows the proportion of total deaths during June to September that we attribute to rainfall shocks. The first row reproduce the result from our main specification that uses PIN code $\times$ year and week $\times$ year (i.e., date) fixed effects (Table 4, Row 1). The second row further interacts the week $\times$ year fixed effects with an indicator representing three areas of Mumbai, which are the same as those used in Figure D.5. The third row interacts the PIN code $\times$ year fixed effects with an indicator for whether the month-of-year is between June and September (inclusive), and also uses week $\times$ year fixed effects. The fourth row uses PIN code $\times$ year, PIN code $\times$ week-of-year, and week $\times$ year fixed effects. Point estimates are calculated based on Equation (7); 95% confidence intervals, calculated with a parametric bootstrap described in Appendix F.1, are in parentheses.