The urban mortality consequences of rainfall and sea level rise

Tom Bearpark* Ashwin Rode† Archana Patankar‡

January 15, 2025

Rainfall and flooding frequently disrupt the lives of urban residents worldwide, posing significant public health risks[1,2]. Rapid urbanisation is exposing larger and more vulnerable populations to these hazards[3], while climate change intensifies rainfall patterns[4], and rising sea levels impair drainage systems[5–9]. Despite the growing recognition and urgency of these hazards, the health impacts of rainfall remain poorly understood, and those of sea level rise are entirely unquantified. Without robust quantification, we lack a complete understanding of the health risks posed by extreme weather and climate change, particularly in rapidly growing megacities where vulnerable populations are disproportionately affected. Here, we estimate the mortality consequences of rainfall in one of the world's largest cities – Mumbai, India. We integrate high-resolution data on rainfall, tides, and mortality, to analyse how extreme rainfall and its interaction with tidal dynamics contribute to urban health risks. We find that rainfall causes roughly 8% of Mumbai's deaths during the monsoon season, and that nearly 90% of this burden is borne by slum-residents. We also find that children face the biggest increase in mortality risk from rainfall, and women face a greater risk than men. Additionally, we demonstrate that mortality risk from rainfall increases sharply during high tides and use this relationship to project how rising sea levels will amplify rainfall-induced mortality in the future. Our findings reveal that the mortality impacts of rainfall are an order of magnitude larger than is documented by official statistics[10], highlighting the urgent need for investment in improved drainage, sanitation, and waste management infrastructure, particularly in cities in the Global South[11]. Moreover, our analysis highlights that the health costs of extreme rainfall and sea level rise are a critical omission in current projections of climate change impacts.

*Princeton University, School of Public and International Affairs.
†University of Chicago, Energy Policy Institute.
‡Green Globe Consulting.
*Correspondence: bearpark@princeton.edu, aarode@uchicago.edu

1 Introduction

Unmanaged rainfall poses a chronic threat to urban areas. Globally, nearly 600 million people are exposed to extreme flood hazard, and this number is projected to grow rapidly due to the urbanisation of flood-prone areas and climate change[3,12]. Unplanned urbanisation, particularly in the Global South, has meant that basic infrastructure, such as drainage, sanitation, and waste management, often fails to keep pace with the growth in built environment and population. As a consequence, rainfall routinely exposes large, densely concentrated populations to damaging and life-threatening flooding hazards. Especially vulnerable are the inhabitants of slum settlements, who number nearly one billion globally[13,14]. Future sea level rise will further amplify flooding hazards[15], making rainwater and urban flood management a defining challenge of 21st century public health, economic development, and urban planning policy.

Despite the growing literature documenting the pervasive and escalating hazards, we still have a very limited understanding of how rainfall and sea level rise translate into health burdens[16]. Estimating these health burdens requires addressing two major challenges. First, since rainfall is often highly localised and its effects unequally distributed, analysing its health impacts demands detailed, within-city data on exposures, vulnerabilities, and health outcomes. Second, statistically disentangling the effects of rainfall from other correlated factors that also influence health is complex. For example, rainfall can affect mortality through a wide range of interconnected pathways, including immediate impacts such as drowning or electrocution, and delayed effects such as disrupted essential services, damaged infrastructure, disease outbreaks, and income shocks[17]. These pathways often overlap, and are confounded by variables like socio-economic conditions, infrastructure quality, and baseline health vulnerabilities, making it difficult to attribute health impacts directly to rainfall exposure.

Faced with these empirical challenges, existing studies focus only on extreme events such as tropical cyclones and analyse their impacts at coarse spatial and temporal resolutions (e.g., annual, state-level)[18–25], leaving the impacts of smaller-scale, frequently occurring, localised rainfall shocks unaccounted for[26]. The lack of sufficiently granular data has also prevented researchers from understanding the differential impacts of rainfall across socio-economic groups within cities. Furthermore, despite the known links between rainfall during high tides, impaired drainage, and urban flooding[5–9], no research to date has examined how future sea level rise might intensify the health consequences of rainfall. As a result, the current body of research provides policy makers with an incomplete and potentially distorted understanding of the health impacts of urbanization, rainfall, climate change, and rising sea levels.

Here, we address both the data and statistical challenges to provide the first population-wide estimates of the all-cause mortality burden of rainfall in one of the world's largest cities – Mumbai, India. We decompose this burden by age, gender, and slum residence, revealing stark differences in vulnerability among urban sub-populations living in close proximity. Additionally, we introduce

a novel empirical approach that leverages the co-occurrence of rainfall and high tides to evaluate how future sea level rise may exacerbate the mortality impacts of rainfall[6–9]. As the majority of the world's megacities are coastal and face rising sea levels[27,28] alongside intensified rainfall[4,29,30], insights from Mumbai offer valuable lessons for addressing health and infrastructure challenges in the world's many rapidly urbanising developing cities such as Manila, Jakarta, and Lagos[27,31–35].

We address the data challenge by constructing a novel dataset that links individual death records to localised, sub-daily rainfall measures during the monsoon season (Methods). These records, collected by Mumbai's governing civic body, the Brihanmumbai Municipal Corporation (BMC), cover the universe of deaths in the city between 2006 and 2015. Each record identifies the residential address of the deceased, which we leverage in two ways. First, we match each death to highly localised rainfall data from the BMC's network of automatic rain gauges, which record rainfall at up to 15-minute intervals across the city. Since ninety percent of Mumbai's population lives within 4 km of a rain gauge (Appendix B.2), our data provide a comprehensive picture of localised rainfall exposure. Second, we classify each deceased individual as either a slum resident or non-slum resident by cross-referencing residential addresses with census lists of slum names and other keywords (Methods).

A simple analysis of the raw data shows clear correlations between rainfall and death in Mumbai. On average, years with more rainfall have more deaths (Fig. 1a), and death rates are substantially elevated during the monsoon season, especially for children under the age of five (Fig. 1b). However, these correlations cannot be interpreted causally as they could be driven by other confounding factors that drive mortality. Our granular dataset, along with the unique meteorology of rainfall in Mumbai, enable us to address the statistical challenge of estimating the causal impact of rainfall on mortality. During the city's June-to-September monsoon season, rainfall on a given day typically occurs in highly localised cloudbursts, delivering heavy rain to some parts of the city while leaving others dry (Fig. 1c, d). The spatial and temporal structure of these cloudbursts is shaped by complex, poorly understood interactions between large-scale weather systems and coastal land-surface features, and has been described as 'arbitrary' and 'patternless,' eluding prediction even by advanced statistical models[36]. As a result, the timing, location, and intensity of rainfall are effectively random within the city on any given day, conditional on the year and date. This natural randomness mirrors the core feature of a randomised controlled trial: random assignment of treatment. It ensures that rainfall 'treatments' are uncorrelated with other factors that might influence mortality, allowing each locality within Mumbai to serve as a valid control before and after receiving rainfall treatment. By leveraging this variation, our 'reduced-form' approach captures the aggregate causal effect of rainfall on mortality across all potential pathways, without requiring detailed enumeration of each mechanism individually[21,37]. Moreover, the effect captured by this approach reflects the net benefits of all adaptive measures undertaken publicly and privately (e.g., water purification, migration in response to flooding).

The mortality impacts of rainfall are mediated by drainage infrastructure, the capacity of which

governs the extent to which rainwater accumulates and translates into flood hazard. An infeasible, and prohibitively expensive, test for the role of this mediating factor might involve experimentally manipulating drainage capacity to investigate whether reduced capacity increases rainfall-driven mortality. However, periodicity in tide heights provides an alternative natural experiment, as high tides are known to impair the carrying capacity of drainage systems[5–9]. By comparing the mortality effects of rainfall that occurs during high vs. low astronomical tide periods of a day we reveal not only the role of drainage today, but also how rising sea levels– by permanently impeding drainage capacity– could substantially increase mortality from rainfall in the future, absent re-engineering.

To measure the effect of rainfall on mortality, we use an econometric model that allows the impact of rainfall on mortality to be non-linear and for the impact to unfold over five-weeks (Methods). To separate the effect of rainfall from other observed and unobserved drivers of mortality, we non-parametrically control for: (1) all average city-wide date-specific patterns and (2) non-linear annual trends specific to each of Mumbai's 89 PIN codes (India's postal area codes, analogous to US ZIP codes). The former accounts for all city level shocks such as terrorist attacks[38,39] and large scale pollution from landfill fires[40], for all seasonal patterns - including environmental changes, and recurring annual events such as holidays, and all long-run non-linear trends such as changing demographics, infrastructure investments, and economic and political conditions. The latter controls for annual trends that vary across PIN-codes, including localised, within-city changes in demographic, infrastructure, and socio-economic conditions.

Our model demonstrates predictive skill (Fig. 2a), capturing much of the rich historical variation in mortality in our data. We conduct extensive checks to ensure that our results are robust to alternative specifications, include those that account for possible spillover effects across PIN codes (Appendix D). Further, we demonstrate that our model does not reflect spurious correlations due to measurement error or model misspecification (Fig. 2d, Methods). We show that our model correctly estimates null results in a series of placebo checks in which we search for spurious associations between our mortality data, and randomly re-shuffled placebo versions of our rainfall data[21,41].

2 Results

2.1 Population-wide impacts

We find that rainfall has a large and persistent impact on all-cause mortality (Fig. 2b). For instance, a single day with 150 mm rainfall causes an increase in 5-week mortality of 2.2% ($\pm$0.65%, $p<0.001$), relative to if that day had been dry. Decomposing this cumulative effect reveals persistent excess mortality over the entire 5-week period. Excess mortality from a rainy day is found to be roughly equal in the contemporaneous week and each of the 4 subsequent weeks (Fig. 2c). This suggests that beyond the immediate mortality repercussions of rainfall in the contempora-

neous week (e.g., drowning, electrocution), rainfall also leads to increased deaths in subsequent weeks as flood-related diseases (e.g., dengue, diarrhoea, malaria, and typhoid), known to peak in the monsoon season (Appendix A), take time to develop and spread[17]. Thus, the persistence of excess mortality over several weeks indicates that rainfall increases the total number of deaths, rather than merely shifting the timing of deaths (i.e., mortality displacement).

2.2 Falsification and robustness checks

We fail to detect an effect of rainfall on mortality when estimating the model with randomly shuffled, false rainfall histories, indicating that our model picks up a real signal in the underlying data[21,41] (Fig. 2d). First, we estimate the model using fully shuffled, randomised rainfall histories, which produces no detectable effect, confirming that our results are not driven by spurious correlations (e.g. due to mismeasured data or an incorrectly specified model). Second, we shuffle rainfall within dates but across PIN codes, which tests whether our findings are driven by broader spatial trends or spillovers rather than localised rainfall effects. Finally, we shuffle rainfall within PIN codes but across dates, ruling out city level trends or seasonality as confounding factors. In each of the checks, our estimated response function from Figure 2b is well outside the envelope of those estimated with the 500 randomly generated relationships. Further, the effects estimated with the reshuffled values are zero on average, with a high degree of precision, indicating that they do not pick up spurious correlations that could be generated by measurement error.

We conduct an extensive suite of robustness checks (Appendix D). We show that our results are nearly identical when estimating the impacts of daily rainfall using daily mortality data rather than weekly mortality data (Appendix D.1). Our results are robust to various functional forms for the causal impact of rainfall on mortality, including those defined non-parametrically over bins of rainfall (Appendix D.2). While our main specifications utilise variation in the daily rainfall, we find similarly large impacts when alternatively investigating the impacts of hourly or weekly rainfall (Appendix D.3). Additionally, we show that our estimates are qualitatively similar when varying the estimating sample (Appendix D.4); the estimation strategy (Poisson, ordinary least squares, and population-weighted regressions; Appendix D.5); when including PIN code-level temperature controls (Appendix D.6); when using different strategies to assign deaths to locations (Appendix D.7); when using alternative non-parametric controls for baseline mortality (fixed effects) (Appendix D.8); and when we make alternative decisions on the processing steps involved in converting our raw station level rainfall data to PIN code-by-day summary statistics (Appendix D.9).

We examine the persistence of mortality effects by estimating regressions with up to 95 lags of daily rainfall (Appendix D.10). Our main model, which limits the effects of rainfall exposure to the past 35 days (5 weeks) likely underestimates the total mortality impact, as effects persist for a longer duration. However, including more lags reduces our sample size and statistical power. We

therefore focus on a 5-week model, acknowledging that it likely understates the persistence of the mortality effect (Discussion). Finally, we pay special attention to the possibility that the impacts of rainfall may extend beyond the PIN code in which it occurs, but do not find strong evidence of such spatial spillovers (Appendix D.10.2).

2.3 Unequal impacts across age groups

Children under the age of 5 are far more vulnerable to the mortality impacts of rainfall than older individuals (Fig. 2a.ii, see also Appendix E.1). A day with rainfall of 150mm causes a 4.7% ($\pm$2.7%, p=0.079) increase in 5-week mortality for the under-5 years age group, but only a 1.6% ($\pm$1.1%, p=0.14) increase for the 5-64 years age group and a 2.4% ($\pm$0.83%, p=0.0042) increase for the 65 years and above age group. These results highlight the welfare importance of rainfall's mortality impacts, as they fall disproportionately on the youngest age group, consistent with the fact that flood-related morbidities such as diarrhoea are known to strongly affect children[42,43]. However, the much higher baseline mortality for older age groups (Fig. 3a.i) means that even small proportional increases due to rainfall translate into a large number of excess deaths among the old (Fig. 5). Nevertheless, the proportional increases by age group that we find contrast sharply with those of prior studies on extreme temperature and mortality, which have primarily found that the oldest age groups are worst affected[37,44].

2.4 Unequal impacts by slum residence

The mortality impacts of rainfall are much larger for individuals residing in slums compared to those not residing in slums (Fig. 3b.ii). A day with 150mm of rainfall causes a 3% ($\pm$0.9%, p<0.001) increase in 5-week mortality for slum residents, but only a 1.1% ($\pm$0.92%, p=0.24) increase for non-slum residents. While the unequal health burden of extreme rainfall across socioeconomic groups has been documented qualitatively before[17,45], our estimates represent the first population-scale causal estimates of such within-city disparity in rainfall-driven mortality impacts. The disparate impacts for slum and non-slum populations exist even when looking only within locations situated at similar elevations, which suggests that simply being located in low-lying, flood-prone areas is not the driving factor behind why slums fare worse during rainfall. Low-elevation PIN codes (i.e., those whose mean elevation is below the median across PIN codes) do suffer higher excess mortality from rainfall than high-elevation PIN codes (Fig. 3b.iii, Appendix E.3). However, in each elevation group, slum populations suffer significantly higher excess mortality than non-slum populations (Fig. 3b.iii). Further, within individual PIN codes, slums are not systematically situated in lower elevation areas (Appendix C.2).

For non-slum dwellers in high elevation PIN codes, the impacts of rainfall are relatively small and statistically insignificant even at high daily rainfall values (Fig. 3b.iii). Thus, similar to[46], we find

that accounting for inequality is crucial 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 adaptative investment.

In Appendix E.2.1 we estimate a model which allows for differential mortality effects of rainfall by age groups within slum and non-slum populations. We find the largest difference in proportional effects across the slum and non-slum populations for the 5-64 years age group, although the effects are imprecisely estimated.

2.5 Unequal impacts by gender

The mortality impacts of rainfall are larger for women than for men (Fig. 3c.ii). A 150mm day causes an increase in 5-week mortality of 1.5% ($\pm$0.86%, p=0.079) for men, but 3.1% ($\pm$0.99%, p=0.0016) for women. While this could be due to physiological factors or social factors (e.g., more resources allocated to males in the event of a weather-induced shock)[47], our data suggest that the age profile of male vs. female deaths plays an important role (Appendix E.2.2). In particular, we find that the female disadvantage is largely driven by non-slum residents in the 65+ age group (Appendix Fig. E.4). Within this particular subpopulation, females tend to die at considerably older ages than their male counterparts (Appendix Fig. E.5), and therefore could be especially vulnerable to the harms brought by rainfall and urban flooding. Our data thus reveal complex patterns of vulnerability to extreme weather among intersections of age, gender, and residence groups within an urban area.

2.6 Subdaily rainfall and tidal height

Our elevation-specific results (Fig. 3b.iii, Appendix E.3) suggest that flooding caused by inadequate drainage is likely a key pathway through which rainfall affects mortality. Another way to understand the role of flooding is to examine subdaily variation in rainfall intensity (Methods). We find that mortality increases sharply with *hourly* rainfall (Fig. 4a). A single hour with 30 mm of rainfall implies an increase in 5-week mortality of 0.6% ($\pm$0.17%, p<0.001). Our results using hourly rainfall are qualitatively consistent with those using daily rainfall, while also revealing that rainfall has higher mortality impacts when it occurs in more intense bursts. For instance, on average, a day with 150 mm of rainfall causes a mortality increase of 2.2% ($\pm$0.65%, p<0.001) (Fig. 2b). If this 150 mm falls over 15 hours with 10 mm in each hour, it leads to a mortality increase of $15 \times 0.13\%$=2% ($\pm$0.79%, p=0.011) (Fig. 4a). However, if this same daily rainfall amount occurs over 5 hours with 30 mm in each hour, it leads to a mortality increase of $5 \times 0.6\%$=3% ($\pm$0.84%, p<0.001).

The coincidence of rainfall and high tides is known to overwhelm drainage systems and increase flood hazard in Mumbai and cities worldwide[5–9] but its health impacts have not been studied

previously. Fig. 4d presents estimates of the mortality impacts of hourly rainfall, allowing the impact to differ based on the tide height during the hour in which the rain falls (e.g., Figs. 4b and c, Methods). We find that rainfall causes relatively little excess mortality when it occurs during a low tide, but that excess mortality is dramatically amplified at high tides. For instance, 30mm of rainfall causes a 1% ($\pm$0.32%, p=0.0018) increase in 5-week mortality if it happens when the predicted tide level is 416 cm (i.e., 95^{th} percentile of the distribution of hourly predicted tide levels in our sample), but only a 0.0074% ($\pm$0.41%, p=0.99) increase if it happens during an hour when the predicted tide level is 94 cm (i.e., 5^{th} percentile) (Fig. 4d). The finding that tide amplifies the impacts of rainfall is robust to alternative ways of modelling the co-occurrence of rainfall and tides, including at the daily level (Appendix E.4).

Since we measure tide using the astronomical tide tables (Methods) rather than observed water levels, the tide is unrelated to rainfall or storm surge, meaning that differences in the mortality effect of rainfall at different tide levels can be interpreted as being caused by the tide itself. We thus find strong evidence that tide is a significant determinant of how rainfall results in excess mortality, pointing to the importance of inadequate drainage and flooding as mechanisms driving the rainfall-mortality relationship.

2.7 The mortality burden of rainfall and sea level rise

Using the estimates from Figs. 2 and 3, we calculate Mumbai's excess mortality from rainfall during our sample period (Methods). This excess mortality amounts to 7.1-8.4% of Mumbai's total deaths in an average monsoon season during 2006 to 2015, depending on the model used, representing 2244-2655 lives lost per year (Fig. 5a, b, Methods, Appendix Fig. F.1). These magnitudes are comparable to the percent of the city's deaths from all cancers[48]. In Appendix F.1 we show that the results are qualitatively robust under a variety of empirical specifications using alternative non-parametric controls for baseline mortality.

Our findings reveal that the excess mortality burden of rainfall is borne unequally across age, gender, and socioeconomic groups. For children under 5, it amounts to 17% of their overall monsoon season deaths— the highest relative proportion among all age groups. However, due to their high baseline mortality, the absolute burden is highest for above 65 year olds, amounting to 12% of overall 65+ monsoon season deaths. Excess mortality is smallest for 5-65 year olds, both in relative and absolute terms (Fig 5a, b). Across genders, relative and absolute excess mortality is substantially larger for women than for men. Furthermore, the excess mortality burden is disproportionately borne by slum residents, representing 11% of monsoon season deaths at slum addresses (1923 deaths per year) compared to 2% at non-slum addresses (292 deaths per year). Thus, we find that 87% of people whose deaths are caused by rainfall are slum-residents.

Applying an India-specific estimate of the value of a statistical life of US$ 1 million[49–51] our results imply that the mortality costs of rainfall in Mumbai are approximately US$2.5 billion per year.

This value only represents the mortality costs; overall welfare impacts would consider many other damages from rainfall and flooding, including morbidity and economic impacts. However, the excess mortality costs of rainfall are likely a substantial, and largely unaccounted, component of total costs. Our estimates of these annual costs for Mumbai alone are comparable in magnitude to India's total annual direct economic damages from flooding[52]. The mortality costs we find for Mumbai are likely to be replicated in urban areas globally, particularly as the cities most vulnerable to flood risk are predominantly in the Global South[27].

High tides amplify the mortality impacts of rainfall (Fig. 4) suggesting a channel through which rising sea levels may contribute to excess mortality. To assess the additional rainfall-driven excess mortality that could occur under higher sea levels, we apply the estimates from Fig. 4c to tide heights that are uniformly increased by the amount of projected sea level rise, while maintaining the historical distribution of rainfall from our data (Methods). This approach simulates how future sea level rise could impede drainage capacity and increase flood hazard from rainfall, just as high tides do so today. It suggests that relative to a baseline case with no sea level rise, sea level rise of 30 cm would increase the excess mortality burden of rainfall from 8.5% to 12% of overall mortality, representing a 37% increase. Sea level rise of 60 cm would increase the burden to 15%– a 77% increase over the no sea level rise baseline. (Fig. 5c). These increases in sea level lie within the range projected for Mumbai by 2100[53], underscoring the significant future harms to human health, absent adaptive infrastructural upgrades such as improved drainage systems.

3 Discussion

This paper has revealed that excess mortality is a major cost of recurrent flooding in an important developing megacity— Mumbai, India, particularly among some of the most vulnerable groups— children under 5, women, and residents of slums. This cost, and its unequal distribution, have not been previously accounted for in analyses of the damages from rainfall or urban flooding. Moreover, our results showing an increased mortality response to rainfall during high tides indicate that future sea level rise could further endanger human lives.

Our findings point to large benefits of protective investments that break the link between rainfall and mortality, especially for the worst affected groups. The ability of water, sanitation, waste management, and man-made and natural drainage systems to cope with rainfall are key challenges common to cities throughout the developing world[54–59]. Investments in safe water and sanitation have been found to drive historical declines in urban child mortality in today's developed countries[60,61]. Our results suggest that similar declines could be achieved in today's developing urban areas, particularly in slums, which bear the vast majority of rainfall's mortality burden (Fig. 3b, Fig. 5). As we enter an era where most of the urban population growth occurs in the developing world, and much of that in slums[14,62], understanding the returns to such investments is a critical area for future research. Climate change is projected to bring additional hazards in the form of

rising sea levels as well as intensified rainfall in many parts of the world[4,29,30], which we show will amplify rainfall's mortality costs, therefore increasing the urgency of intervention.

Several caveats to our analysis are worth noting. First, the rainfall data make it difficult to capture mortality impacts that may occur over very long time horizons (i.e., longer than 5 weeks), because the measurements of local rainfall are only available for a portion of each year (Appendix B.2). Accounting for these longer term effects may result in larger impact estimates (Appendix D.10.1), given the extended duration that certain rainfall-related diseases can take to develop (e.g. malaria[63]).

Second, while our mortality data captures all-cause mortality, it does not report the cause of death, precluding us from fully understanding the exact pathways through which rainfall imperils human health across urban subpopulations. Nevertheless, aggregate statistics suggest that rainfall-related diseases are an important pathway, as the seasonal increase in deaths during Mumbai's monsoon (Fig. 1b) is mirrored in the increased incidence of several diseases known to be linked to urban flooding— dengue, diarrhoea, malaria and typhoid (Appendix A)[17]. The continued impacts over five weeks following a rainy day (Fig. 2c) indicate that such diseases are likely to be a key factor, beyond immediate causes such as drowning or electrocution. Moreover, the large impacts that we find for children under 5 accord with the fact that flood-related morbidities such as diarrhoea are known to strongly affect children[42,43].

Finally, our findings that excess mortality impacts of rainfall are stronger in low elevation localities and during periods of high tide suggest that flooding is a key channel for the impacts of rainfall. However, we do not directly observe flood depth or duration across locations following rainfall events. Flooding data would enable further understanding of how rainfall accumulation propagates across space and time and harms human health, and also the value of specific improvements in drainage infrastructure. Flood observations will also be a valuable input into flood modelling, both today as well as for studying the implications of future sea level rise in a more nuanced manner. Although our simplified counterfactual analysis is likely appropriate for the range of sea level rise we consider (Fig. 5c), more extreme sea level rise may alter tides in more complex ways and require more sophisticated modelling[64].

While these limitations point to important areas for future research, we demonstrate here the feasibility of evaluating the health impacts of unmanaged rainfall, both today and under future sea level rise, by combining granular data on rainfall, tides, and vital statistics. Our approach is particularly useful given the scarcity of more complex data (e.g., flood maps, drainage infrastructure maps) in developing country settings[65,66], and the difficulties of flood detection by satellite in densely built urban areas[67–69]. Future studies could replicate our approach in other cities and integrate this new category of impacts into global climate change damage estimates. Recent work has recognised the importance of rainfall in projecting the GDP implications of climate change[25,70]. However our findings on mortality suggest that GDP alone is insufficient to capture the full scope of climate-related impacts. The emphasis placed on temperature-related effects in climate impact assessments[37,71]

should be expanded to incorporate the critical role of rainfall and sea level rise, particularly in the vulnerable urban environments in which a large and growing proportion of the world lives.

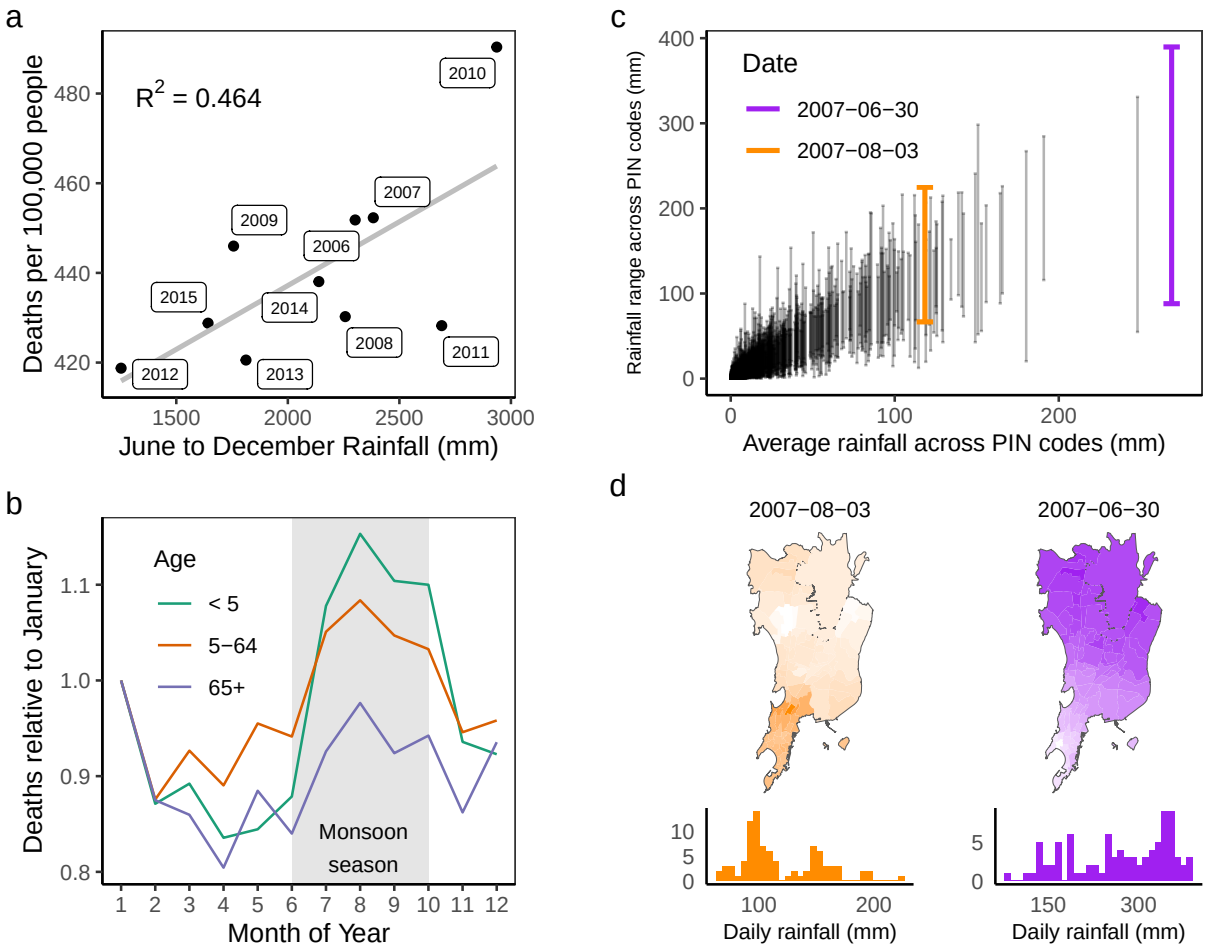


Figure 1: **Rainfall and death in Mumbai, India: Temporal and spatial variation.**
a, Positive correlation between total June to December annual deaths per 100,000 people and total June to December rainfall ($R^2 = 0.464$). Each point represents an individual year during our 2006-2015 sample period. **b**, Ratio of 2006-2015 average age-wise mortality rate by month of year, relative to January. Shaded area denotes monsoon season months (June to September). **c**, Spatial variation in daily rainfall across locations in Mumbai, illustrating localised intensity of monsoon rainfall. Each bar represents an individual day in the sample period. Bars vertically range from the minimum to maximum rainfall total across PIN codes (postal area codes) on that day, and are indexed by the day's average rainfall total across PIN codes (horizontal axis). Rainfall on the coloured dates is mapped in d). **d**, Maps showing spatial patterns of rainfall totals across Mumbai's PIN codes on two illustrative days (2007-08-03 and 2007-06-30), with darker shades indicating higher rainfall. Histograms show distribution of daily rainfall totals across PIN codes for the corresponding maps.

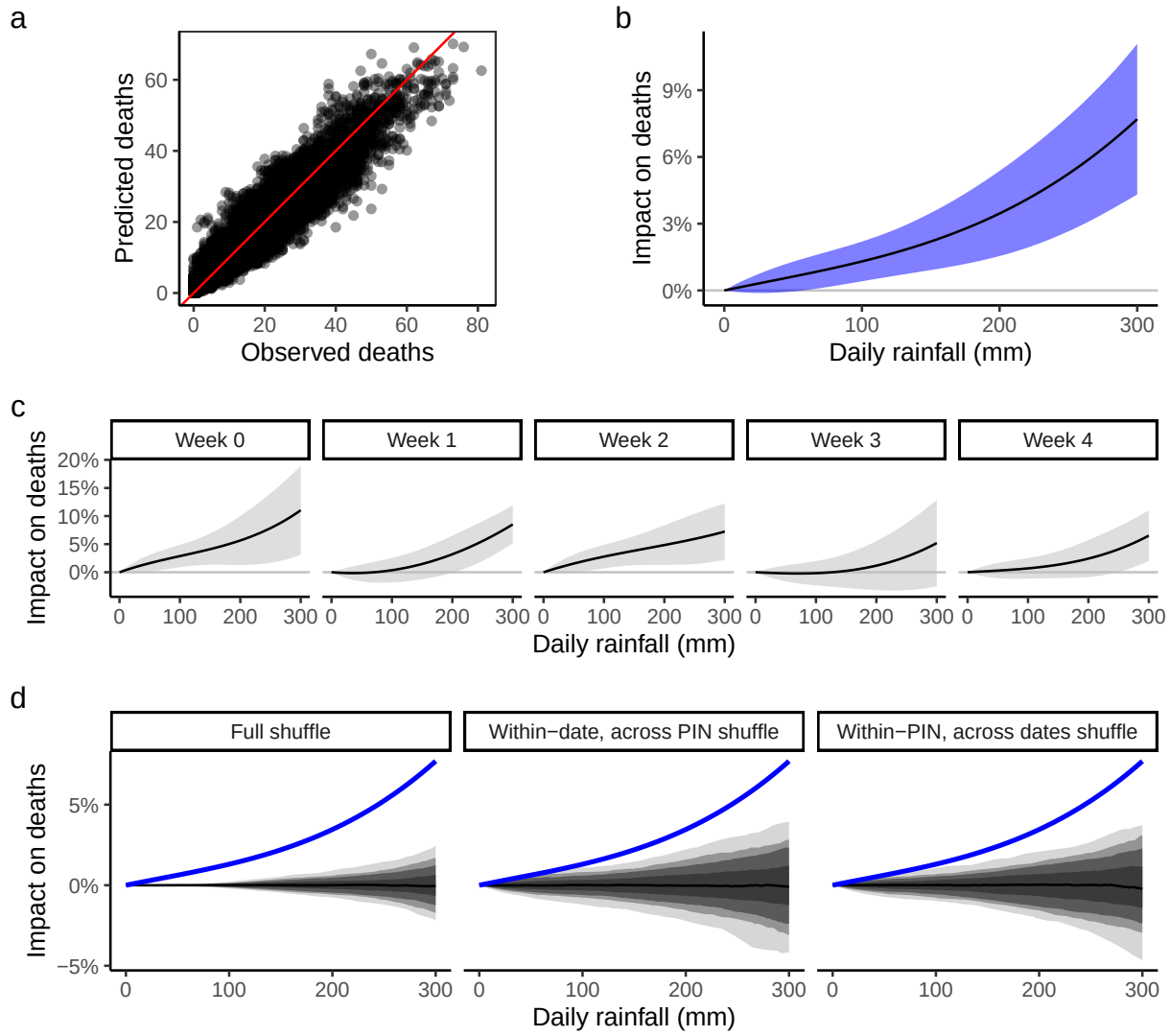


Figure 2: **The causal effect of daily rainfall on mortality.**
a, Predicted versus observed deaths from our empirical model. Each point represents a PIN code × week. Diagonal line denotes 45°. **b**, The casual effect of rainfall on deaths. Points along the curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall (Methods). Shaded areas indicate 95% confidence intervals. Regression results are reported in tabular form in Appendix C.1. **c**, Decomposes the effect shown in **a**. 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. **d**, Placebo randomization tests to validate the estimated rainfall-death relationship. Each panel represents different shuffling schemes for rainfall. The leftmost panel shows estimates using rainfall values that have been shuffled randomly across both dates and PIN codes. The middle panel shows estimates using rainfall values that have been shuffled randomly across PIN codes, but within the same date. The rightmost panel shows estimates using rainfall values that have been shuffled randomly across dates, but within the same PIN codes. The blue lines reproduce our central estimate from **a** that uses the true rainfall data. In each panel, shaded areas show the 1st, 5th, 10th, 25th 75th, 90th, 95th, and 99th percentiles of the distribution across 500 randomised estimates using shuffled rainfall, while the solid black line denotes the median of the randomised estimates.

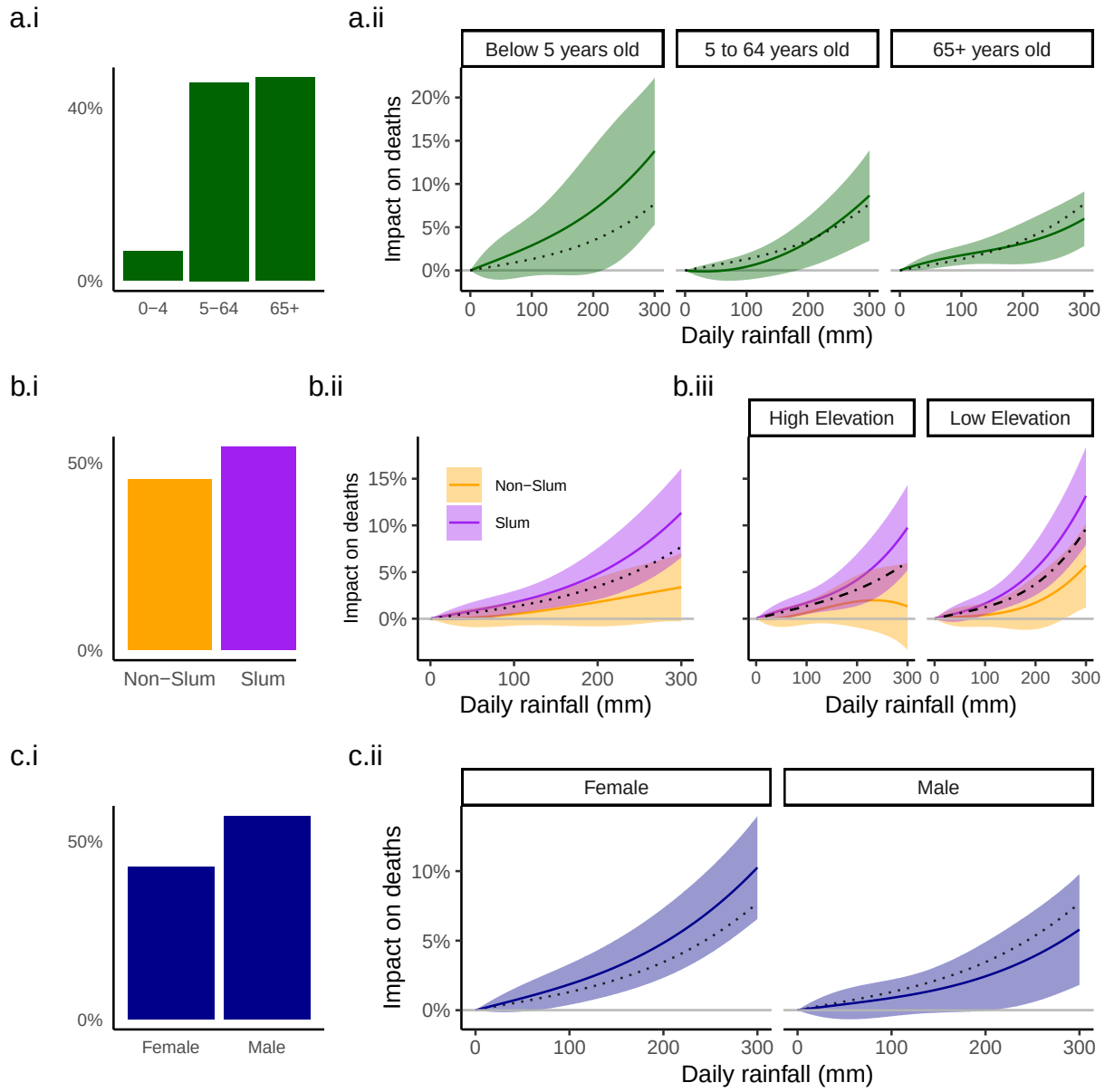


Figure 3: **Unequal mortality consequences of rainfall across population subgroups.** Throughout, points along each curve represent the percent change in 5-week mortality due to a single day with the total rainfall value shown on the x-axis, relative to a day with zero rainfall. Dotted black line reproduces the population-wide average mortality effect of daily rainfall shown in Fig. 2. Regression results are reported in tabular form in Appendix C.1. **a.i**, Age distribution of deaths in our estimating sample. **a.ii**, Impact of daily rainfall on deaths by age group. **b.i** Percentage of deaths from slum and non-slum addresses in our estimating sample. **b.ii**, Impact of daily rainfall on deaths for slum and non-slum populations. **b.iii**, Impact of daily rainfall on deaths for populations at low versus high elevations (dashed lines), and slum and non-slum populations at low and high elevations (solid lines). **c.i**, Gender distribution of deaths in our estimating sample. **c.ii**, Impact of daily rainfall on death rates by gender.

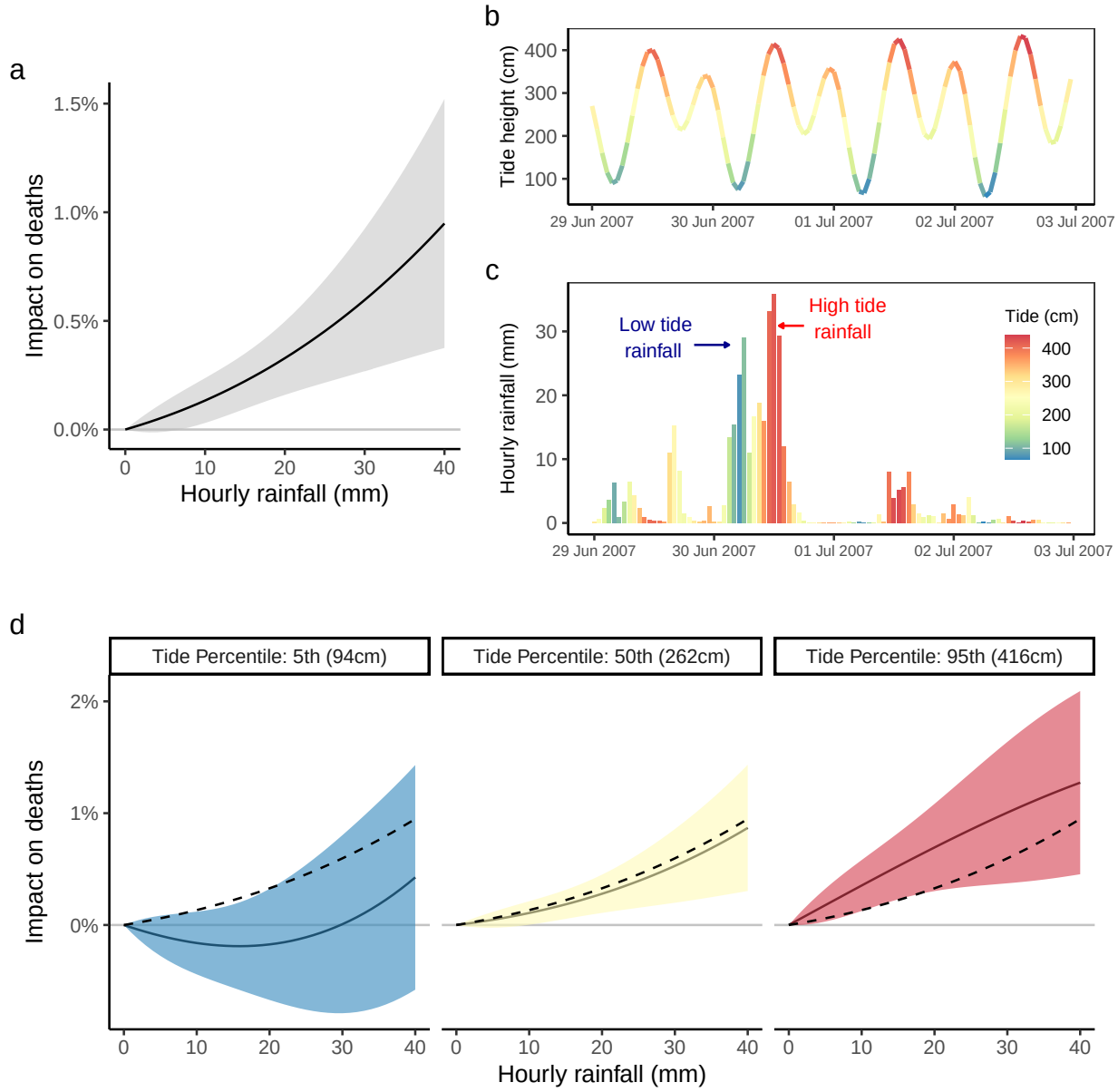


Figure 4: **Amplified mortality impact of rainfall during high tides.**
a, Impact of hourly rainfall on deaths. Points along the curve represent the percent change in 5-week mortality due to a single hour with the total rainfall value shown on the x-axis, relative to an hour with zero rainfall (Methods). **b**, Tide height variations over selected dates (29 June to 3 July, 2007), highlighting cyclical high and low tides within each day. **c**, Hourly rainfall over the same dates, colour-coded by tide height during which the rainfall occurs. **d**, Impact of hourly rainfall on deaths, when rainfall occurs during a low tide height (5th percentile in sample, blue), median tide height (yellow), or a high tide height (95th percentile, red). Dashed lines reproduces baseline estimate from **a**. Regression results are reported in tabular form in Appendix C.1 for the response functions presented in **a** and **d**.

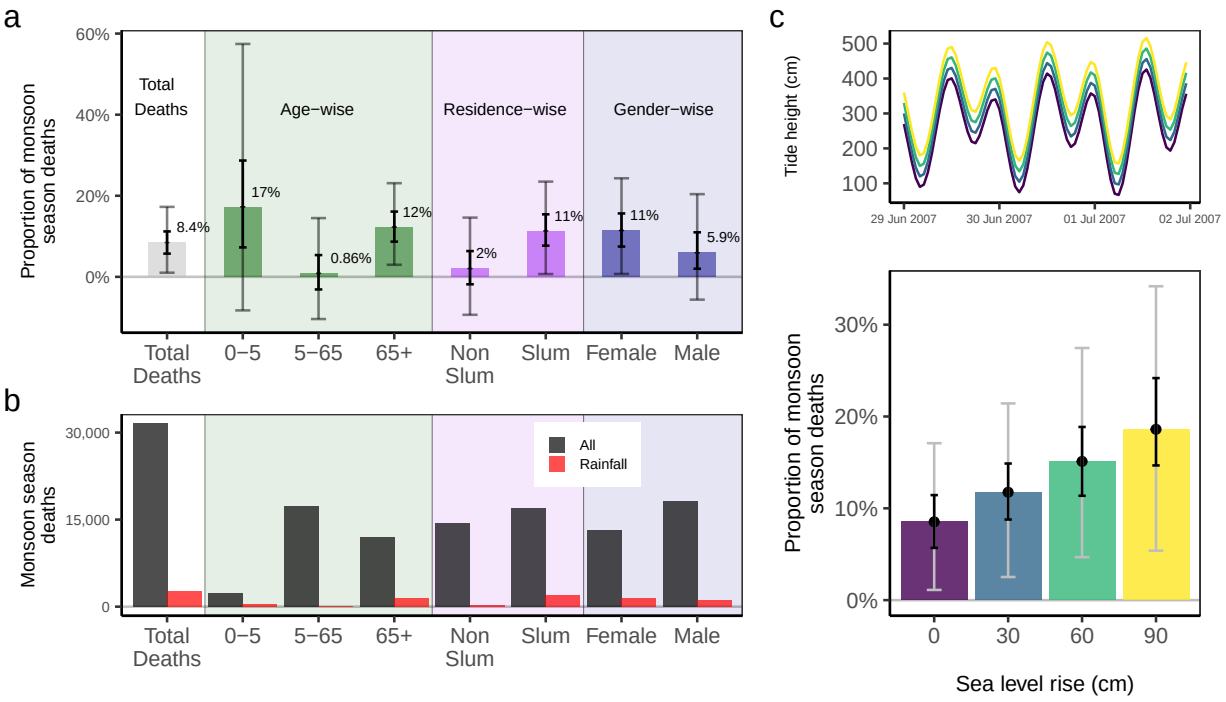


Figure 5: **The mortality burden of rainfall and future sea level rise.**
a, The proportion of monsoon season deaths caused by rainfall. Bar shows excess mortality from rainfall as a proportion of overall mortality for the full population (left-most bar, based on Fig. 2b estimates) and by age, residence, and gender (respectively based on Fig. 3a.ii and 3b.ii, 3c.ii estimates). Grey error bars show 95% confidence intervals, and the black error bars show the interquartile range (Appendix F.2). See Appendix F.1 for these results in tabular format. **b**, Annual deaths caused by rainfall (red bars), compared to total deaths (black bars), based on proportions shown in **a**. **c**, Excess mortality from sea level rise. Top panel illustrates tide heights over selected dates (29 June to 3 July, 2007) elevated by 0 cm, and 30 cm, 60 cm, and 90 m, which correspond to future sea level rise scenarios within the range of projections for Mumbai[53]. Bottom panel shows excess monsoon season mortality from rainfall as proportion of overall mortality under these sea level rise amounts. Excess mortality is calculated based on estimates from Fig. 4d by uniformly increasing tide heights by the values indicated on the x-axis (Methods). As a baseline, leftmost bar illustrates the mortality burden of rainfall with no sea level rise.

Methods

Here we provide an overview of the data sources, econometric model, and calculation of the mortality burden of rainfall today and under future sea level rise. Further details on these can be found in the Appendix.

Superscripted numbers in blue are used to indicate footnotes, while superscripted numbers in black are used to indicate references.

Data sources

All-cause mortality data

Our mortality data are obtained through systematic queries of the Mumbai municipal government's deaths registration system website.[1] The resulting data provide us with the universe of recorded deaths in Mumbai between 2001 and 2015.[2] In our analysis, we use the deaths from 2006 to 2015 because granular weather data 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 government reports 94,706 in its official aggregate statistics.[3] Thus, in 2015, our data 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.[4] Additionally, We use the residential address text 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

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

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

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

[4] 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.1.1. In Appendix D.7, we conduct robustness checks where we use the coarser, ward-level location identifier and retain observations lacking a PIN code.

terms that are clearly associated with slum addresses (Appendix B.1.2). Appendix B summarises the spatial and temporal patterns in the mortality data, including inequalities within the city across slum and non-slum communities.

Weather data

Our main weather data are rainfall measurements collected from a dense network of Automatic Weather Stations (AWS), which were set up by Mumbai's municipal government after record-setting floods in July 2005[5]. 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.[5] The AWS data also include temperature measurements, which we use as a control variable in a robustness check (Appendix D.6). 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[72]. 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.2. The data are summarised in Fig. 1 and Appendix B.2.

Other data

We supplement the 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 $\times$ 50 m, using the R package elevatr[6], 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.[7]

To calculate hourly tide levels for Mumbai, we use historical tide tables from the US National Oceanographic and Atmospheric Administration.[8] These provide us with the high and low tide predictions for each day in our sample. We only have one time series of tide predictions available to us, which is for 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

[5] 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.2. In Appendix D.9 we show robustness of our main results to creating our climate data measures using only a balanced sample of stations that are operating in all years.

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

[7] In an additional robustness check (Appendix E.3), 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 HAND map for Mumbai developed in ref.[73], which we obtained through direct correspondence.

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

Mumbai's coastline[74]. 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 (Fig. 4).

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 and rainfall from the Indian Meteorological Department's (IMD) two weather stations in Mumbai, and use these to validate the AWS data (Appendix B.2). We obtained a shapefile[75] for the locations of slums in Mumbai through correspondence with the authors, which we use to characterise the elevation of slum areas within a PIN code (Appendix C.3). Finally, we construct PIN code-level population estimates, using the Global Human Settlement Layer (GHSL) satellite-derived data product[76], 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.5).

We provide summary statistics for the data used in our main regression analysis in Appendix B.3.

Econometric model

Nonlinear effect of daily rainfall on weekly mortality

We flexibly model PIN code-level weekly deaths as a function of the PIN code's daily rainfall totals within the contemporaneous and preceding weeks. Importantly, our econometric procedure is designed to recover non-linear changes in deaths in response to daily rainfall. Allowing for such daily non-linearity is critical, because different days in a week may have very different rainfall amounts, generating potentially disparate effects on mortality[77].

Our estimating equation takes the following form:

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

where i indexes the PIN code, w indexes the week of the year, and y indexes the calendar year. The outcome variable $m_{i,w,y}$, which is the expected number of log deaths in a PIN code $\times$ week, is modelled as a linear-in-parameters function f of a vector of rainfall variables, $\boldsymbol{R}^{\text{daily}}_{i,w,y}$, and a granular set of location- and time-specific intercepts, referred to here as "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[78].

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 rainfall within the week[79].[9] Second, the arguments of f include lagged-week values of the vector of rainfall variables. This is important because a day's rainfall in a given week may affect mortality in subsequent weeks. For instance, an increase in mortality due to an extreme rainfall day in a given week could be compensated for by a fall in mortality in subsequent weeks. This could occur if the extreme rainfall causes a deterioration in the health of individuals who would have died even in its absence. In this case, the only effect of the rainfall is to alter the timing of mortality by a few weeks, but not the number of deaths in the longer run.[10] Alternatively, an extreme rainfall day 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 has persistent effects that result in a higher number of deaths in the longer run. Including lagged rainfall values in f accommodates both these possible narratives[81].

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.[11] 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:

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

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

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,[12] while the date fixed effects flexibly control for any city-wide shocks in a given week as well as within-year seasonality. The use of fixed effects accounts for location-specific, time-trending, and cyclical determinants of mortality non-parametrically and is thus more reliable than trying to explicitly enumerate and control for the complex and unobserved individual determinants of

[9] For instance, the effect of a week with 280 mm of total rainfall will differ if all the 280 mm fell on a single day vs. if 40 mm fell on each day of the week.

[10] This type of temporal displacement is sometimes referred to as a "harvesting" effect[80].

[11] We emphasise results from the third-order polynomial because it captures important non-linearities while being relatively parsimonious. We emphasise 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 D, we show results using alternative functional forms (Appendix $D.1$) and number of lags (Appendix $D.10.1$).

[12] 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. In Appendix D.9, we show that our results are qualitatively unchanged when only using a subset of stations that operate in all years.

mortality. Together, our fixed effects imply that we identify the effect of rainfall on mortality from idiosyncratic spatial variation in rainfall within the city in a given time period, allowing for a causal interpretation of the function f. This research design is possible due to the highly localised nature of where rain falls on a given day in Mumbai (Fig. 1c, d). Fig. 2 summarises the results from estimating Equation 1.

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 present a series of randomisation placebo checks to show that spurious correlations (e.g., due to mismeasured rainfall and/or mortality) are not driving our results (Fig 2d) [21,41]. In Appendix D, we present an extensive suite of robustness and sensitivity checks related to various assumptions in processing the data and estimating the econometric model.

Effects by age, gender, and residence

We model the differential effect of rainfall on mortality based on individual-level characteristics by estimating a separate response function for individuals in each group a:

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

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, $a \in \{< 5 \text{ years}, [5, 65) \text{ years}, \geq 65 \text{ years}\}$ (Appendix E.1 considers alternative age categories), and we classify each individual's residence as slum or non-slum based on residential address (See Appendix B.1.2). Fig. 3 summarises the results from estimating Equation 3 for various groups.

Within-day timing of rainfall

While Equations 1 and 3 model the non-linear effect of *daily* rainfall on weekly mortality, we also consider a model where PIN code-level weekly deaths are a function of the PIN code's *hourly* rainfall distribution within the contemporaneous and preceding weeks. This econometric procedure recovers non-linear changes in deaths in response to hourly rainfall. The estimating equation takes the following form:

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

The expected number of log deaths in a PIN code $\times$ week is modelled as a linear-in-parameters function g of a vector of rainfall variables, $\boldsymbol{R}^{\text{hourly}}_{i,w,y}$, and fixed effects. The function g here causally relates hourly rainfall to mortality outcomes, allowing for non-linearity in the response of weekly mortality to hourly rainfall. To construct it, we take non-linear transformations (i.e., polynomials) of hourly rainfall before summing these over days of the week. The function g thus captures the impacts of the distribution of rainfall across hours within the week[79].[13]

We construct g using a third-order polynomial in hourly rainfall, for days in the contemporaneous week and 4 lagged weeks. Letting $R^k_{i,h,d,w,y}$ represent the total rainfall on hour h of day d raised to the power k:

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

where each parameter, $\beta^{hourly}_{l,k}$, denotes a coefficient on the k^{th} power and l^{th} weekly lag term. Figure 4a summarises the results from estimating Equation 4.

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

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

We consider two additional models that leverage variation in astronomical tide patterns (Appendix E.4), both of which find that high tides increase rainfall damages. First, we consider a model in which hourly rainfall is interacted with an indicator for whether the predicted tide is above or below 450cm. A broad range of Mumbai municipal government documents, NGO reports, and news alerts state a threshold tide level of 450 cm, over which acute flooding is very likely to occur.[14] Second, we consider a model in which we interact the daily average predicted tide level with the daily rainfall response function f from Equation (2).

[13] For instance, the effect of a week with 280 mm of total rainfall will differ if the rain fell over 10 hours with 28 mm in each hour vs. if the rain fell over 40 hours with 7 mm in each hour.

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

Calculating the mortality burden of rainfall

The mortality-rainfall responses that we estimate in Equations 1, 3, 4, and 6 can be used to calculate the excess mortality from rainfall during our estimation period[21].

Using the population-wide response (Equation (1)), the log of expected deaths given a rainfall vector $\boldsymbol{R}^{\text{daily}}_{i,w,y}$ is $f(\boldsymbol{R}^{\text{daily}}_{i,w,y})+\alpha_{i,y}+\lambda_{w,y}$. These deaths consist of those due to idiosyncratic rainfall, 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 is thus $\alpha_{i,y} + \lambda_{w,y}$. After converting logs to levels, we estimate the expected number of deaths caused by rainfall, $r_{i,w,y}$, by subtracting total deaths from those not caused by rainfall:

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

We use our empirical estimates of all parameters that appear 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 (Figure 5b), which can be expressed as a proportion of total observed monsoon season in-sample deaths (Figure 5a).[15]

This exercise can be similarly implemented for other models, including those with mortality-rainfall responses by age group, slum residence, and gender (Equation 3). For instance, in the case of responses that differ 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. 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 monsoon season in-sample deaths in age category a.

Sea level rise and the mortality burden of rainfall

To calculate the rainfall-driven excess mortality that would occur under a higher sea level, we use the tide-differentiated model (Equation 6). Let $g_{\tau+\Delta}(\boldsymbol{R}^{\text{hourly}}_{i,w,y})$ denote the mortality-rainfall response function given a sea level rise of Δ. This function is constructed by uniformly increasing tide heights in $g_{\tau}(\boldsymbol{R}^{\text{hourly}}_{i,w,y})$ by Δ:

$$g_{\tau+\Delta}(\boldsymbol{R}^{\text{hourly}}_{i,w,y}) = g(\boldsymbol{R}^{\text{hourly}}_{i,w,y}) + \sum_{k=1}^{3}\sum_{l=0}^{4} \gamma^{\text{hourly}}_{l,k} \sum_{d \in w-l} \sum_{h \in d} \left(R^{k}_{i,h,d,w-l,y} \cdot (\tau_{h,d,w-l,y} + \Delta) \right). \tag{8}$$

[15] 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.2.

Following the approach of Equation 7, we estimate $r_{i,w,y,\Delta}$, the expected number of deaths caused by rainfall given that the sea level increases the predicted tides by Δ:

$$r_{i,w,y,\Delta} = \exp(\alpha_{i,y} + \lambda_{w,y}) \left(\exp(g_{\tau+\Delta}(\boldsymbol{R}_{i,w,y}^{\text{hourly}})) - 1 \right). \tag{9}$$

Summing the estimate, $\hat{r}_{i,w,y,\Delta}$, over the tide height and rainfall distribution in all PIN codes and weeks in our sample gives us an estimate of the total in-sample mortality burden of rainfall if the sea level was elevated by Δ.[16] In Figure 5c, we express this as a proportion of total observed monsoon season in-sample deaths.

[16] It should be noted that modelling the effect of a sea level rise of Δ as a uniform increase in tide heights by Δ may not be valid for extreme sea level rise scenarios (e.g., $\Delta \sim 10$ m) but is a reasonable representation for the scenarios we consider here (i.e., $\Delta < 1$ m) [64].

Acknowledgements

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 Subrata Kumar Das, Miki Doan, Subimal Ghosh, Simon Greenhill, Michael Greenstone, Dylan Hogan, Amir Jina, Sulakshana Mahajan, Yogesh Mishra, Michael Oppenheimer, Tom Scott, Yixin Sun, Vaidehi Tandel, Shrabani Tripathy, Guglielmo Zappalà, and workshop participants at the World Bank, Stanford University, National Bureau of Economic Research, 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.

References

[1] Jha, A. K., Bloch, R. & Lamond, J. *Cities and flooding: a guide to integrated urban flood risk management for the 21st century* (World Bank Publications, 2012).

[2] Guan, M., Liang, Q. & Hou, J. Smart approaches to predict urban flooding: Current advances and challenges. *Frontiers in Earth Science* **9**, 681751 (2021). Publisher: Frontiers Media SA.

[3] Rentschler, J. *et al.* Global evidence of rapid urban growth in flood zones since 1985. *Nature* **622**, 87–92 (2023). URL `https://www.nature.com/articles/s41586-023-06468-9`. Number: 7981 Publisher: Nature Publishing Group.

[4] Field, C. B., Barros, V., Stocker, T. F. & Dahe, Q. *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change* (Cambridge University Press, 2012).

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

[6] Zellou, B. & Rahali, H. Assessment of the joint impact of extreme rainfall and storm surge on the risk of flooding in a coastal area. *Journal of Hydrology* **569**, 647–665 (2019). Publisher: Elsevier.

[7] Dong, L. *et al.* The influence of astronomical tide phases on urban flooding during rainstorms: Application to Macau. *Journal of Hydrology: Regional Studies* **56**, 101998 (2024). Publisher: Elsevier.

[8] Zhuang, Y., Xu, K., Bin, L., Wang, C. & Shen, R. Assessment of the jacking effect of high tide on compound flooding in a coastal city under sea level rise based on water tracer modeling. *Journal of Hydrology* **627**, 130474 (2023). Publisher: Elsevier.

[9] Qiang, Y. *et al.* Urban flood analysis for Pearl River Delta cities using an equivalent drainage method upon combined rainfall-high tide-storm surge events. *Journal of Hydrology* **597**, 126293 (2021). Publisher: Elsevier.

[10] Central Water Commission and others. State-wise data on damage caused due to floods during 1953–2011 (2012). Publisher: Government of India New Delhi, India.

[11] Bryan, G., Glaeser, E. & Tsivanidis, N. Cities in the developing world. *Annual Review of Economics* **12**, 273–297 (2020). URL `https://www.annualreviews.org/content/journals/10.1146/annurev-economics-080218-030303`.

[12] Tellman, B. *et al.* Satellite imaging reveals increased proportion of population exposed to floods. *Nature* **596**, 80–86 (2021). URL `https://www.nature.com/articles/s41586-021-03695-w`. Number: 7870 Publisher: Nature Publishing Group.

[13] Handmer, J. *et al.* Changes in impacts of climate extremes: human systems and ecosystems. In *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the Intergovernmental Panel on Climate Change*, 231–290 (Cambridge University Press (CUP), 2012).

[14] Marx, B., Stoker, T. & Suri, T. The economics of slums in the developing world. *Journal of Economic perspectives* **27**, 187–210 (2013). Publisher: American Economic Association.

[15] Pariartha, I. P. G. S. *et al.* Compounding effects of urbanization, climate change and sea-level rise on monetary projections of flood damage. *Journal of Hydrology* **620**, 129535 (2023). URL https://www.sciencedirect.com/science/article/pii/S0022169423004778.

[16] Hsiang, S. Global economic impacts of climate change: an empirical perspective. *Annual review of economics* .

[17] Patankar, A. M. The exposure, vulnerability, and ability to respond of poor households to recurrent floods in Mumbai. *World Bank Policy Research Working Paper* **1** (2015).

[18] Anttila-Hughes, J. & Hsiang, S. Destruction, disinvestment, and death: Economic and human losses following environmental disaster. *Available at SSRN 2220501* (2013).

[19] Kocornik-Mina, A., McDermott, T. K. J., Michaels, G. & Rauch, F. Flooded Cities. *American Economic Journal: Applied Economics* **12**, 35–66 (2020). URL https://www.aeaweb.org/articles?id=10.1257/app.20170066.

[20] Gandhi, S., Kahn, M. E., Kochhar, R., Lall, S. & Tandel, V. Adapting to Flood Risk: Evidence from a Panel of Global Cities (2022). URL https://papers.ssrn.com/abstract=4134933.

[21] Young, R. & Hsiang, S. Mortality caused by tropical cyclones in the United States. *Nature* **635**, 121–128 (2024). URL https://www.nature.com/articles/s41586-024-07945-5. Publisher: Nature Publishing Group.

[22] Elliott, R. J., Strobl, E. & Sun, P. The local impact of typhoons on economic activity in China: A view from outer space. *Journal of Urban Economics* **88**, 50–66 (2015). Publisher: Elsevier.

[23] Holtermann, L. Precipitation anomalies, economic production, and the role of "first-nature" and "second-nature" geographies: A disaggregated analysis in high-income countries. *Global Environmental Change* **65**, 102167 (2020). URL https://www.sciencedirect.com/science/article/pii/S0959378020307500.

[24] Damania, R., Desbureaux, S. & Zaveri, E. Does rainfall matter for economic growth? Evidence from global sub-national data (1990–2014). *Journal of Environmental Economics and Management* **102**, 102335 (2020). URL https://www.sciencedirect.com/science/article/pii/S0095069620300589.

[25] Kotz, M., Levermann, A. & Wenz, L. The effect of rainfall changes on economic production. *Nature* **601**, 223–227 (2022). URL https://www.nature.com/articles/s41586-021-04283-8. Number: 7892 Publisher: Nature Publishing Group.

[26] 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.

[27] Hallegatte, S., Green, C., Nicholls, R. J. & Corfee-Morlot, J. Future flood losses in major coastal cities. *Nature Climate Change* **3**, 802–806 (2013). URL https://doi.org/10.1038/nclimate1979.

[28] Rosenzweig, C. *et al.* The future we don't want: How climate change could impact the world's greatest cities. *C40 Cities Secretariat: London, UK* (2018).

[29] Wang, B. *et al.* Monsoons Climate Change Assessment. *Bulletin of the American Meteorological Society* **102**, E1–E19 (2021). URL https://journals.ametsoc.org/view/journals/bams/102/1/BAMS-D-19-0335.1.xml. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[30] Salunke, P., Keshri, N. P., Mishra, S. K. & Dash, S. K. Future projections of seasonal temperature and precipitation for India. *Frontiers in Climate* **5** (2023). URL https://www.frontiersin.org/articles/10.3389/fclim.2023.1069994.

[31] Rentschler, J., Braese, J., Jones, N. & Avner, P. Three feet under: The impact of floods on urban jobs, connectivity, and infrastructure. *World Bank Group Policy Research Working Paper* (2019). Publisher: World Bank Group.

[32] Stoutjesdijk, J. Philippines-Metro Manila Flood Management Project: environmental and social impact assessment (Vol. 2): Environmental and social impact assessment: executive summary (2017). URL https://policycommons.net/artifacts/1281166/philippines-metro-manila-flood-management-project/1873600/. Publisher: < bound method Organization. get_name_with_acronym of< Organization: World ....

[33] Dasgupta, S. *et al. Urban Flooding of Greater Dhaka in a Changing Climate: Building local resilience to disaster risk* (World Bank Publications, 2015). URL https://books.google.nl/books?hl=en&lr=&id=X2XmCgAAQBAJ&oi=fnd&pg=PP1&dq=Urban+flooding+of+Greater+Dhaka+in+a+changing+climate+:+building+local+resilience+to+disaster+risk+(English)&ots=OdFLo_6PDv&sig=Tyv1YgVnkqYj7tJMa80Wh6ylIqY.

[34] Higuera Roa, O. *et al.* Technical Report: Lagos floods. Tech. Rep., United Nations University - Institute for Environment and Human Security (2022). URL http://collections.unu.edu/view/UNU:9024. Publisher: United Nations University-Institute for Environment and Human Security (UNU-EHS).

[35] Hsiao, A. Sea Level Rise and Urban Adaptation in Jakarta (2023).

[36] Singh, J. *et al.* Spatio-temporal analysis of sub-hourly rainfall over Mumbai, India: Is statistical forecasting futile? *Journal of Earth System Science* **126**, 38 (2017). URL https://ui.adsabs.harvard.edu/abs/2017JESS..126...38S. ADS Bibcode: 2017JESS..126...38S.

[37] Carleton, T. *et al.* Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits*. *The Quarterly Journal of Economics* **137**, 2037–2105 (2022). URL https://doi.org/10.1093/qje/qjac020.

[38] Times, N. Y. Series of Bombs Explode on 7 Trains in India, Killing Scores (2006). URL https://www.nytimes.com/2006/07/12/world/asia/12india.html.

[39] CNN. Mumbai Terror Attacks Fast Facts (2022). URL https://www.cnn.com/2013/09/18/world/asia/mumbai-terror-attacks/index.html.

[40] Times, L. A. 18 stories high and still burning, fire at landfill exposes India's growing trash crisis (2016).

[41] Hsiang, S. M. & Jina, A. S. The causal effect of environmental catastrophe on long-run economic growth: Evidence from 6,700 cyclones. Tech. Rep., National Bureau of Economic Research (2014).

[42] Kosek, M., Bern, C. & Guerrant, R. L. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bulletin of the world health organization* **81**, 197–204 (2003). Publisher: SciELO Public Health.

[43] Escobar Carias, M. S., Johnston, D. W., Knott, R. & Sweeney, R. Flood disasters and health among the urban poor. *Health Economics* **31**, 2072–2089 (2022). Publisher: Wiley Online Library.

[44] Heutel, G., Miller, N. H. & Molitor, D. Adaptation and the Mortality Effects of Temperature across U.S. Climate Regions. *The Review of Economics and Statistics* **103**, 740–753 (2021).

[45] WRI-India. Climate and Air Pollution Risks and Vulnerability Assessment for Mumbai, India. Tech. Rep., WRI (2022).

[46] Lindersson, S. *et al.* The wider the gap between rich and poor the higher the flood mortality. *Nature Sustainability* **6**, 995–1005 (2023). URL https://www.nature.com/articles/s41893-023-01107-7. Number: 8 Publisher: Nature Publishing Group.

[47] Fruttero, A. *et al. Gendered Impacts of Climate Change: Evidence from Weather Shocks*. Policy Research Working Papers (The World Bank, 2023). URL http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-10442.

[48] Praja Foundation. Cause of death in mumbai: 2012 to 2020 (2020). Place: Mumbai.

[49] Shanmugam, K. R. Discount Rate for Health Benefits and the Value of Life in India. Working Paper, Madras School of Economics, Chennai, India (2011). URL https://econpapers.repec.org/paper/madwpaper/2011-060.htm.

[50] Majumder, A. & Madheswaran, S. Value of statistical life in India: A hedonic wage approach. Working Paper 407, Institute for Social and Economic Change, Bangalore (2018). URL https://econpapers.repec.org/paper/schwpaper/407.htm.

[51] Sweis, N. J. Revisiting the value of a statistical life: an international approach during COVID-19. *Risk Management* **24**, 259–272 (2022). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9017085/.

[52] WMO. State of the Climate in Asia 2022. Tech. Rep., World Meteorological Organization (WMO) (2022). URL https://library.wmo.int/records/item/66314-state-of-the-climate-in-asia-2022.

[53] Garner, G. *et al.* IPCC AR6 WGI sea level projections. Tech. Rep., IPCC (2022). Publisher: World Data Center for Climate (WDCC) at DKRZ.

[54] Tingsanchali, T. Urban flood disaster management. *Procedia engineering* **32**, 25–37 (2012). Publisher: Elsevier.

[55] Marsalek, J. *Urban water cycle processes and interactions: urban water series-UNESCO-IHP* (CRC press, 2014).

[56] Lamond, J., Bhattacharya, N. & Bloch, R. The role of solid waste management as a response to urban flood risk in developing countries, a case study analysis. *WIT Transactions on Ecology and the Environment* **159**, 193–204 (2012). Publisher: WIT Press.

[57] Tucci, C. E. Urban flood management. *WMO and Capnet* (2007).

[58] Silveira, A. d. Problems of modern urban drainage in developing countries. *Water Science and Technology* **45**, 31–40 (2002). Publisher: IWA Publishing.

[59] Parkinson, J. & Mark, O. *Urban stormwater management in developing countries* (IWA publishing, 2005).

[60] Alsan, M. & Goldin, C. Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920. *Journal of Political Economy* **127**, 586–638 (2019). Publisher: The University of Chicago Press Chicago, IL.

[61] Anderson, D. M., Charles, K. K. & Rees, D. I. Reexamining the contribution of public health efforts to the decline in urban mortality. *American Economic Journal: Applied Economics* **14**, 126–157 (2022). Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

[62] UN DESA. World urbanization prospects: The 2018 revision (2018).

[63] Liang, S. Y. & Messenger, N. Infectious diseases after hydrologic disasters. *Emergency medicine clinics of North America* **36**, 835–851 (2018). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6195322/.

[64] Pickering, M. *et al.* The impact of future sea-level rise on the global tides. *Continental Shelf Research* **142**, 50–68 (2017). URL https://linkinghub.elsevier.com/retrieve/pii/S0278434316304824.

[65] See, L. S., Calo, L., Bannon, B. & Opdyke, A. An open data approach to mapping urban drainage infrastructure in developing communities. *Water* **12**, 1880 (2020). Publisher: MDPI.

[66] Nkwunonwo, U., Whitworth, M. & Baily, B. A review of the current status of flood modelling for urban flood risk management in the developing countries. *Scientific African* **7**, e00269 (2020). Publisher: Elsevier.

[67] Islam, M. T. & Meng, Q. An exploratory study of Sentinel-1 SAR for rapid urban flood mapping on Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation* **113**, 103002 (2022). Publisher: Elsevier.

[68] Schumann, G., Giustarini, L., Tarpanelli, A., Jarihani, B. & Martinis, S. Flood modeling and prediction using earth observation data. *Surveys in Geophysics* **44**, 1553–1578 (2023). Publisher: Springer.

[69] Tanim, A. H., McRae, C. B., Tavakol-Davani, H. & Goharian, E. Flood detection in urban areas using satellite imagery and machine learning. *Water* **14**, 1140 (2022). Publisher: MDPI.

[70] Bilal, A. & Känzig, D. R. The Macroeconomic Impact of Climate Change: Global vs. Local Temperature (2024). URL https://www.nber.org/papers/w32450.

[71] Rennert, K. *et al.* Comprehensive evidence implies a higher social cost of CO2. *Nature* **610**, 687–692 (2022). URL https://www.nature.com/articles/s41586-022-05224-9. Publisher: Nature Publishing Group.

[72] Auffhammer, M., Hsiang, S. M., Schlenker, W. & Sobel, A. Using Weather Data and Climate Model Output in Economic Analyses of Climate Change. *Review of Environmental Economics and Policy* **7**, 181–198 (2013). URL https://doi.org/10.1093/reep/ret016. Publisher: The University of Chicago Press.

[73] Tripathy, S. S., Vittal, H., Karmakar, S. & Ghosh, S. Flood risk forecasting at weather to medium range incorporating weather model, topography, socio-economic information and land use exposure. *Advances in Water Resources* **146**, 103785 (2020). URL https://www.sciencedirect.com/science/article/pii/S0309170820304164.

[74] National Institute of Oceanography. Study of Land Reclamation and its Influence on Mumbai's Coastline and Implications for Proposed Projects. Tech. Rep., Mumbai Transformation Support Unit (2016).

[75] Tandel, V., Gandhi, S., Patranabis, S., Bettencourt, L. M. A. & Malani, A. Infrastructure, enforcement, and COVID-19 in Mumbai slums: A first look. *Journal of Regional Science* **62**, 645–669 (2022). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jors.12552. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jors.12552.

[76] Schiavina, M. *et al.* GHSL Data Package 2023 (2023).

[77] Rode, A. *et al.* Estimating a social cost of carbon for global energy consumption. *Nature* **598**, 308–314 (2021). URL https://www.nature.com/articles/s41586-021-03883-8. Publisher: Nature Publishing Group.

[78] Chen, J. & Roth, J. Logs with Zeros? Some Problems and Solutions. *The Quarterly Journal of Economics* **139**, 891–936 (2024). Publisher: President and Fellows of Harvard College.

[79] Hsiang, S. Climate Econometrics. *Annual Review of Resource Economics* **8**, 43–75 (2016). URL https://doi.org/10.1146/annurev-resource-100815-095343. Publisher: Annual Reviews.

[80] Deschênes, O. & Moretti, E. Extreme Weather Events, Mortality, and Migration. *The Review of Economics and Statistics* **91**, 659–681 (2009). Publisher: MIT Press.

[81] Deschênes, O. & Greenstone, M. Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US. *American Economic Journal: Applied Economics* **3**, 152–185 (2011). URL https://www.aeaweb.org/articles?id=10.1257/app.3.4.152.

[82] World Economic Forum. These are the world's five biggest slums (2020). URL https://www.weforum.org/agenda/2016/10/these-are-the-worlds-five-biggest-slums/.

APPENDIX

Table of contents below provides page numbers for both main text and Appendix sections.

Superscripted numbers in blue are used to indicate footnotes, while superscripted numbers in black are used to indicate references.

Contents

1 Introduction **2**

2 Results **4**

3 Discussion **9**

A Seasonality in disease incidence **33**

B Data **34**

C Regression Tables **48**

D Robustness checks **51**

E Differential effects of rainfall on mortality by age, gender, residence, elevation, and tide height during rainfall **64**

F The mortality burden of rainfall: Alternative models and confidence interval calculation **72**

A Seasonality in disease incidence

This appendix investigates the seasonality in disease incidence across Mumbai. We obtain monthly data (2008-2019) on disease cases reported at each of the medical dispensaries run by Mumbai's governing civic body, the Brihanmumbai Municipal Corporation (BMC), through correspondence with a non-governmental organisation, the Praja Foundation.[17] Figure A.1 plots time series (top row) and monthly means (bottom row) of these data, for four diseases— dengue, diarrhea, malaria, and typhoid— that are known to be triggered by urban flooding[17]. All four diseases show clear signs of elevated incidence during monsoon months.

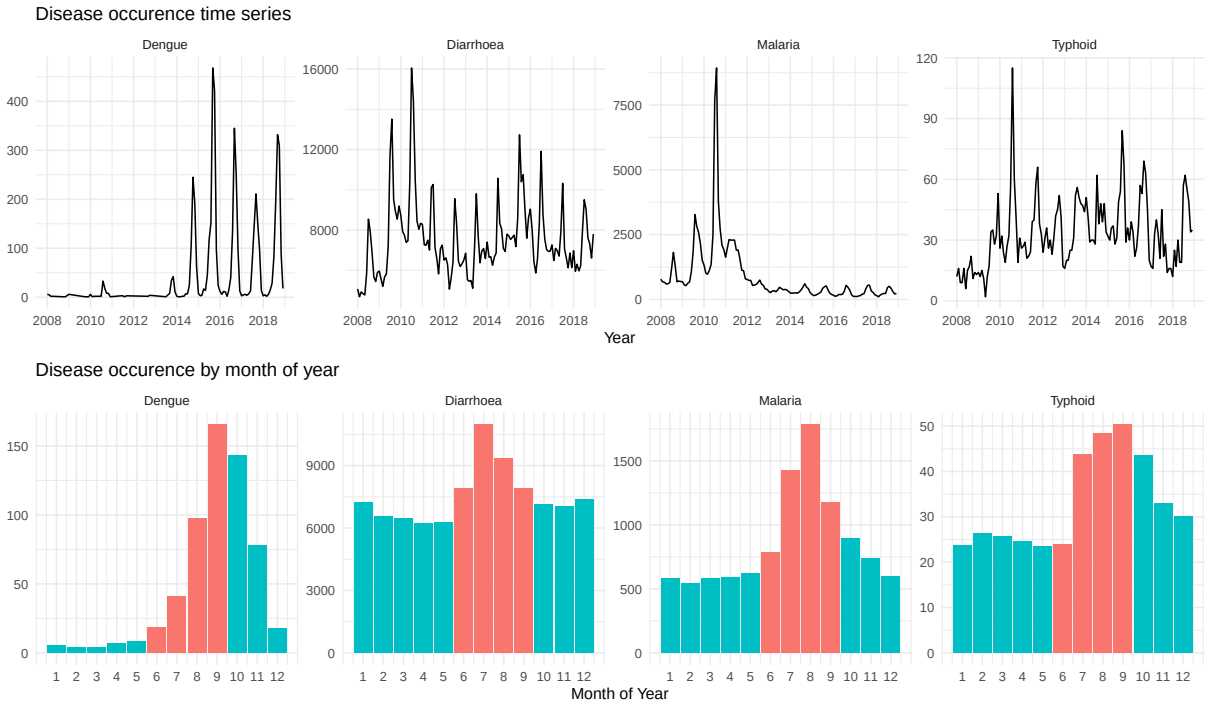


Figure A.1: Seasonality of disease incidence
Top row shows monthly time series of the city-wide case count for four selected diseases across all municipal dispensaries between 2008 and 2019. Bottom row shows monthly means across these years to illustrate seasonality. Monsoon season months are coloured in red. Data are obtained from the Praja Foundation, who collected the raw data through a Right To Information (RTI) request.

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

B Data

This appendix describes further details of the mortality data (Section B.1) and weather data (Section B.2) used in our analysis, including the procedures for processing of each dataset. The appendix concludes with a summary of the variables used to estimate the econometric model (Section B.3).

B.1 Mortality data

Fig. B.1 illustrates the stark spatial patterns in the data. Panel A shows the mean age of death in our mortality data between 2006 and 2015, mapped across Mumbai's 89 PIN codes (i.e., postal area codes, equivalent to US ZIP 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 47 and 40 respectively.[18] 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.

The deaths data also reveal variation and inequities along other dimensions.[19] 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.[20] Furthermore, an increasing trend in the mean age of death is found in 87 out of the 89 PIN codes.[21]

[18] Dharavi is widely regarded as one of the world's largest slums[82].

[19] 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.

[20] 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`.)

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

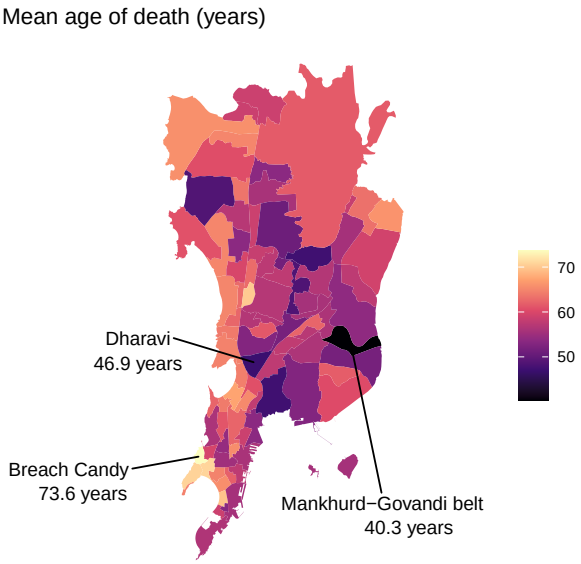


Figure B.1: Spatial patterns in mortality in Mumbai.
Figure shows the spatial distribution of the mean age of death between 2006 and 2015 across PIN codes in Mumbai. We highlight the PIN codes containing two of Mumbai's biggest slum areas, Dharavi (400017) and the Mankhurd-Govandi belt (400043), as well as the affluent Breach Candy neighbourhood (400026).

B.1.1 Location assignment in mortality data

We assign deaths to locations using one of two approaches. In the approach used in our main results, we assign deaths to PIN codes using the PIN code from the residential addresses of the deceased provided in the mortality records. However, this approach requires us to throw out roughly 20% of our observations, for which the address does not record a PIN code. Therefore in an alternative approach, we assign deaths to each of Mumbai's 24 municipal wards, using the ward indicated in each mortality record. Sensitivity of our results to this alternative approach is shown in Appendix D.7.

B.1.2 Slum classification

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

First, we obtain a list of 1034 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.[22] 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

[22]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.

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.[23] 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 deaths 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 (Figure B.2). Second, at the PIN code level, we compare the mean age of death in the slum addresses to that in the non-slum addresses (Figure B.3). 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 outside slums.

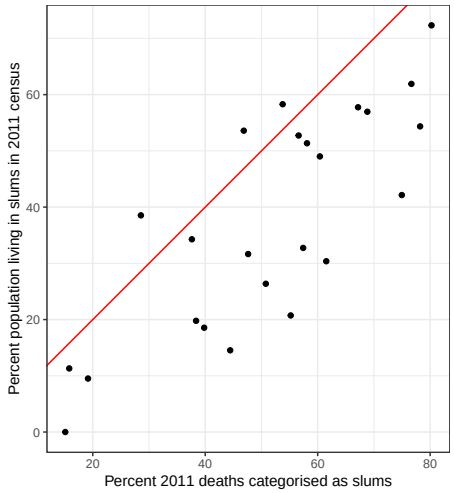


Figure B.2: Comparison of slum proportion categorisation to census data
The 2011 census provides the proportion of residents of each municipal ward that live in slums. Plot compares these census values to our categorisation of the proportion of deaths in each ward from slum addresses. The red line shows a 45 degree line. The R^2 value for this relationship is 0.63.

[23] 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`.

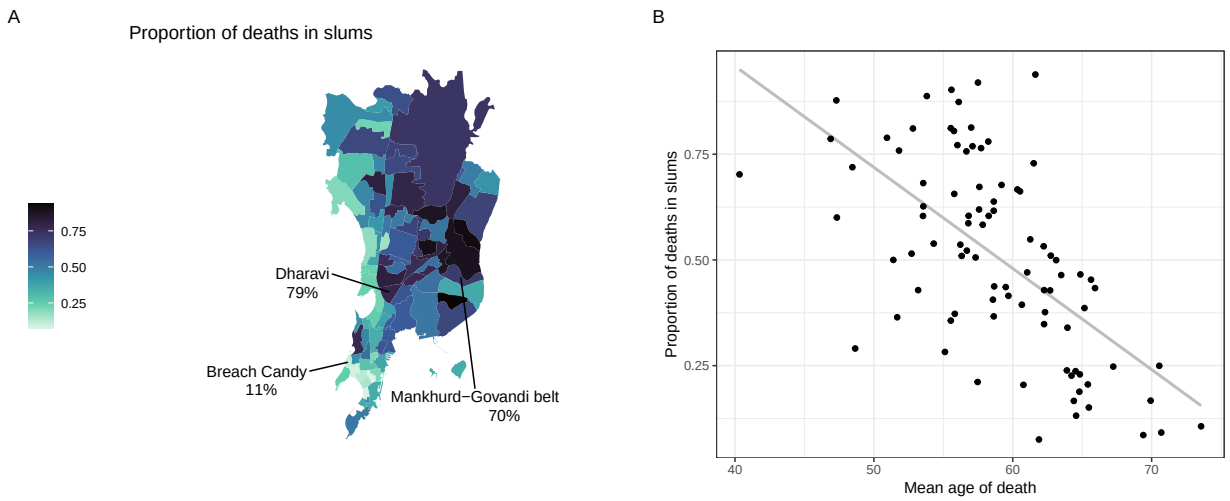


Figure B.3: Proportion of deaths in slums and mean age of death, by PIN code.
Map shows the spatial distribution of the percent of 2006-2015 deaths in slums across PIN codes in Mumbai, highlighting the PIN codes containing two of Mumbai's biggest slum areas— Dharavi (400017) and the Mankhurd-Govandi 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.

B.2 Weather data

We obtain weather data at fifteen minute intervals from Mumbai's governing civic body, the Brihanmumbai Municipal Corporation (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 the BMC following record-setting rainfall and flooding in July 2005[36].

Figure B.4 shows the locations of the set of stations we have available to us in each year. Throughout the sample period, the network is extremely dense, and provides good geographical coverage across the city. This is important, since Mumbai's rainfall is highly localised. Over all sample years, the average distance to a station is less than 2 km for a resident of Mumbai, and this value is generally stable across time and similar across slum and non-slum populations (Figure B.5). Moreover, ninety percent of Mumbai's population lived within 4 km of a rain gauge (Figure B.6).



Figure B.4: 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. Colors indicate whether stations were available in all 10 years of our estimating sample.



Figure B.5: Distance to a weather station for Mumbai residents
Mean distance to a weather station for slum and non-slum residents of Mumbai by year, based on 100 m $\times$ 100 m 2010 gridded population data from Global Human Settlement Layer[76].

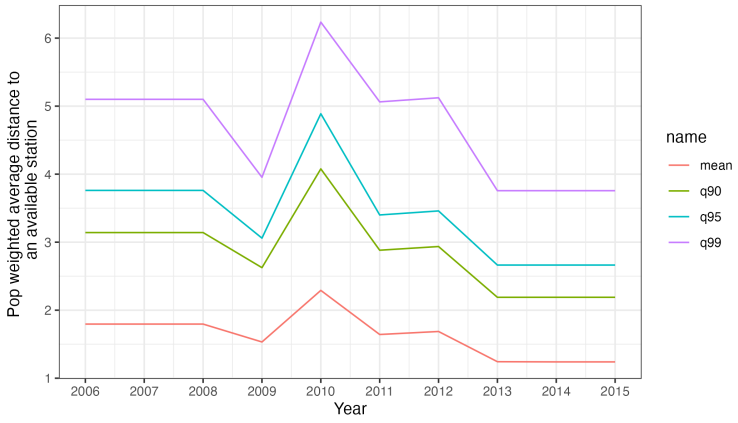


Figure B.6: Distance to a weather station for Mumbai residents
Distance to a weather station for residents of Mumbai by year: mean, and 90th, 95th, and 99th percentiles, based on 100 m $\times$ 100 m 2010 gridded population data from Global Human Settlement Layer[76].

B.2.1 Processing weather station data

The weather data required considerable processing to prepare it for regression analysis. We next outline the steps taken to process the raw station level data into PIN code level variables to be combined with our mortality data.

Step 1: Cleaning station names and dates

First, we match station names in the 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 following data cleaning steps:

- We drop station-days in each year where there are fewer than half the stations available.
- We linearly interpolate short spells of missing data. In particular, if a station-day is missing a spell of 1 hour or less of rainfall, or a spell 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 (i.e., both temperature and rainfall readings) 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 or rainfall values at the 15-minute by station level with NAs. We define implausible values as:

- Temperature above 42.2° C, or below 10° C (i.e., the maximum and minimum recorded 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[24] 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.
  - Overall, we end up imputing around 9% of the station-15 minute period temperature values, and around 5% of the rainfall values.

Step 3: Creating zonal statistics

Step 2 provided us with a panel of station-15 minute level rainfall and temperature data, that is balanced within each year, but not across years.

Our next step is to create PIN code and municipal ward (i.e., 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 hourly or daily level

We next sum our 15 minute level data to the daily or hourly levels.

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 (Methods Equation (2)) or hourly (Methods Equation (5)) level, in order to preserve important non-linearities[79]. These include binning or polynomials.

Step 6: Temporal aggregation

Finally, we aggregate the transformed data sets from step 5 to the weekly level that is used in our main text regressions by taking a simple summation (Methods Equations (2) and (5)).

[24] 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.

B.2.2 Diagnostics of weather data

In order to ensure they are capturing plausible variation, we conduct a series of diagnostic checks on the station level and PIN code level rainfall data. First, we examine the pair-wise correlation of daily rainfall in every station pair. Figure B.7 plots correlation coefficients of daily rainfall values in each year for all possible pairings of stations, revealing that rainfall in nearby station pairs exhibits higher positive correlation than does rainfall in more distant station pairs. Second, Table B.1 demonstrates that despite high spatial variation in where rain falls on individual days (Figure 1c,d), both the average and 99^{th} percentile daily rainfall values over the full estimating sample are very similar across PIN codes and are consistent with the region's climatology. Third, we compare the temporal variation in rainfall and temperature to that from the Indian Meteorological Department's (IMD) weather station at Mumbai's airport (Figure B.8). Overall, the temporal patterns in the AWS data are similar to those in the IMD data.

Finally, we conduct falsification tests by estimating our econometric model with various random shufflings of the rainfall data (Methods). We uncover a positive relationship between rainfall and mortality only when using the true rainfall data and not the reshuffled data (main text Figure 2d), strongly suggesting that the AWS data picks up real variation in rainfall.

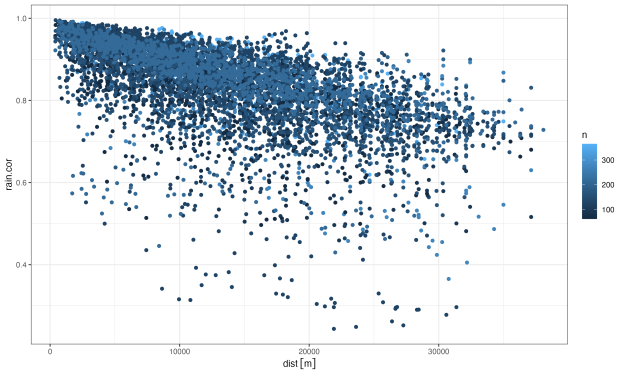


Figure B.7: Spatial correlation of daily station level rainfall values
Plot shows the correlation between all station pairs of daily rainfall values, where each point is the correlation between a pair of station in a given year. Vertical axis shows the correlation coefficient while horizontal axis shows the distance in meters between the stations in the pair. Colors represent the number of observations used to calculate each correlation.

Table B.1: Rainfall distribution in estimating sample by PIN code
Table shows the PIN code level summary statistics of the daily rainfall distribution across days in the estimating sample. The 'Avg.' column shows the average daily rainfall, while $q99$ shows the 99^{th} percentile of daily rainfall. Finally, the 'Days above 100mm' column denotes the number of days with at least 100mm of rainfall.

PIN code	Avg.	q99	Days $>$ 100mm	PIN code	Avg.	q99	Days $>$ 100mm
400001	10.93	126.75	39	400057	11.31	129.15	35
400002	10.41	127.13	34	400058	11.51	136.39	42
400003	10.95	128.94	42	400059	11.36	129.00	46
400004	10.54	123.93	35	400060	11.48	124.38	41
400005	10.38	124.32	37	400061	11.38	121.82	42
400006	10.37	125.04	34	400063	11.44	126.60	38

400007	10.07	124.68	36	400064	10.41	129.88	38
400008	10.83	128.22	43	400065	11.45	122.17	39
400009	11.11	131.44	43	400066	11.72	122.09	38
400010	10.93	127.66	42	400067	11.16	124.89	37
400011	10.88	129.11	43	400068	12.45	141.37	48
400012	11.04	128.83	43	400069	11.50	135.50	40
400013	10.90	125.74	41	400070	12.53	145.85	48
400014	10.94	131.33	45	400071	11.71	133.83	47
400015	11.12	128.00	41	400072	11.87	132.04	47
400016	11.34	132.70	46	400074	11.67	133.87	45
400017	11.42	134.44	47	400075	12.05	134.07	47
400018	10.87	127.06	40	400076	12.03	128.88	49
400019	11.53	130.82	46	400077	12.28	142.42	48
400020	10.41	124.14	37	400078	11.53	124.63	44
400021	10.28	124.32	36	400079	12.04	133.42	46
400022	11.57	132.93	43	400080	12.92	142.95	43
400024	11.84	134.59	47	400081	12.90	134.18	48
400025	10.53	123.81	43	400082	12.73	141.76	43
400026	10.35	122.85	34	400083	12.03	125.59	48
400027	10.97	136.62	44	400084	12.33	140.43	50
400028	11.11	127.72	43	400085	11.60	130.23	44
400029	11.72	135.68	44	400086	12.17	139.17	47
400030	10.49	122.77	43	400087	11.68	125.06	46
400031	11.14	129.25	44	400088	11.81	129.73	44
400032	10.34	125.84	38	400089	12.15	136.51	46
400033	11.02	129.70	42	400091	11.00	122.62	36
400034	10.53	123.81	35	400092	11.34	120.14	38
400035	10.55	124.72	37	400093	11.48	128.70	43
400037	11.44	135.32	42	400094	11.74	131.55	46
400042	12.05	124.82	44	400095	11.24	126.93	36
400043	11.94	132.48	47	400096	11.50	126.71	41
400049	11.34	128.61	39	400097	11.39	123.73	34
400050	11.16	134.73	43	400098	11.49	138.46	45
400051	11.45	136.33	46	400099	11.73	132.26	46
400052	11.31	130.62	43	400101	11.33	124.71	36
400053	11.46	128.69	43	400102	11.51	132.08	43
400054	11.32	129.39	37	400103	11.79	131.49	42
400055	11.42	131.04	38	400104	11.38	128.48	39
400056	11.32	131.64	36	NA	NA	NA	NA

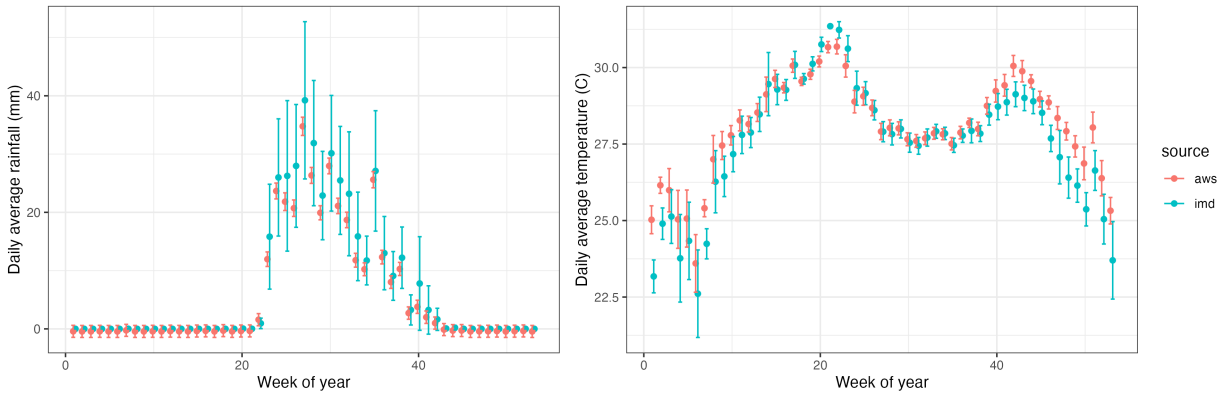


Figure B.8: AWS and IMD comparison

Plot compares average daily temperature and precipitation values by week of year in the AWS data with those in the Indian Meteorological Department's (IMD) data from its weather station at Mumbai's airport. AWS data are averaged across PIN-codes, days within each week of year, and years in the sample. IMD data are averaged across days within each week of year and years in the same sample as the AWS data. Error bars represent 95% confidence interval for the mean values.

B.3 Estimation data summary

B.3.1 Dates of estimation sample

The availability of AWS data determines the dates of our estimation sample. These data are available when the stations are in operation, which is usually during the monsoon season months in each year. Table B.2 describes the temporal extent of the sample in each year and the number of deaths by year during the sample period.[25] Table B.3 presents the same information, but broken down by week of year.

[25] Specifically, the deaths data that is part of our estimating sample must be in a week during which there exists AWS data for that week as well as in the number of lagged weeks included in the model (e.g., 4 lagged weeks as specified in Methods Equation 1).

Year	Total deaths in raw data	Total deaths in cleaned data	Dates in estimating sample	Total deaths in estimating sample
2006	89661	67729	2006-07-23 to 2006-12-24	31254
2007	91011	73917	2007-01-01 to 2007-12-24	73695
2008	90127	73746	2008-01-01 to 2008-09-23	55559
2009	90650	72870	2009-02-05 to 2009-09-24	47119
2010	97010	76870	2010-07-02 to 2010-10-22	28889
2011	89390	70177	2011-07-02 to 2011-11-05	26362
2012	88925	69161	2012-07-01 to 2012-08-19	11265
2013	90078	69389	2013-07-02 to 2013-07-23	5328
2014	92447	70980	2014-07-02 to 2014-10-22	24285
2015	91797	69718	2015-07-02 to 2015-09-24	17496

Table B.2: Weeks in estimating sample by year

Table traces the number of deaths in the raw mortality data to the number of deaths in our estimating sample by 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, which are mainly available only during the monsoon season in each year (see also Table B.3).[26]

[26] 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 traces the number of deaths in our raw data 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.

B.3.2 Summary statistics of estimation variables

Table B.4 presents summary statistics for the variables used in estimating the econometric models described in Methods.

	Mean	SD	Median	Min	q^{99}	Max	Total Deaths
Total Deaths	15.8	11.5	13	0	51	128	363380
Deaths below 5 yrs	1.1	1.9	0	0	8	26	24853
Deaths 5-64 yrs	7.2	6.1	6	0	27	105	165710
Deaths 65 yrs and above	7.4	5.4	7	0	23	39	169674
Male deaths	9.0	7.0	8	0	31	107	207593
Female deaths	6.8	5.3	6	0	23	36	155744
Slum deaths	8.6	8.6	6	0	37	63	197577
Non-Slum deaths	7.2	5.9	6	0	25	93	165803
Daily rainfall total	11.4	26.4	0.6	0	130.7	389.7	
Weekly rainfall total	80.3	120.2	23.3	0	520.4	766.4	

Table B.4: 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} and standard deviations by SD. Totals deaths summed across age groups 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 B.5). We exploit this variation in rainfall across PIN codes at the daily level to estimate a non-linear effect of daily rainfall on mortality.

[27] This is the average during our estimation sample, which is usually only the monsoon season (Section B.3.1).

Year	N	$\geq$ 50mm	$\geq$ 100mm	$\geq$ 150mm	$\geq$ 200mm	$\geq$ 300mm
2006	17177	1174	542	184	35	0
2007	32485	1168	390	170	67	40
2008	24386	1443	317	60	1	0
2009	24297	818	375	137	52	0
2010	13528	1741	391	37	3	0
2011	14952	1215	578	171	11	0
2012	8188	529	87	4	0	0
2013	5340	1091	420	94	12	0
2014	13528	1271	362	32	0	0
2015	10769	765	272	89	75	13

Table B.5: Extreme rainfall observations by year.
Table shows the total number of PIN code $\times$ days (N) observed in each year in the estimating sample, 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 dates of the sample in each year are in Section B.3.1.

C Regression Tables

This appendix displays the mortality-rainfall response estimates from main text Figures 2, 3, and 4 in tabular form (Section C.1). In addition, it displays estimates from a regression of elevation on slum status using grid cell level data (Section C.2).

C.1 Mortality-rainfall response estimates

	(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	0.0062	0.014	-8e-04	0.01	0.0043	0.0087	0.0011	0.0084
	(0.0034)	(0.013)	(0.0053)	(0.0043)	(0.0055)	(0.0049)	(0.0052)	(0.0047)
100	0.013	0.029	0.0043	0.018	0.0088	0.019	0.0049	0.017
	(0.0045)	(0.018)	(0.0077)	(0.0057)	(0.0067)	(0.0075)	(0.0064)	(0.0063)
150	0.022	0.047	0.016	0.024	0.015	0.031	0.011	0.03
	(0.0065)	(0.027)	(0.011)	(0.0083)	(0.0086)	(0.0099)	(0.0092)	(0.009)
250	0.052	0.1	0.057	0.043	0.038	0.072	0.026	0.075
	(0.013)	(0.043)	(0.02)	(0.015)	(0.017)	(0.016)	(0.016)	(0.02)
300	0.077	0.14	0.087	0.06	0.058	0.1	0.034	0.11
	(0.017)	(0.043)	(0.027)	(0.016)	(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 2b. Columns 2 to 4 show the response functions from main text Figure 3a.ii (estimated for 3 age groups in a stacked regression). Columns 5 and 6 show results from main text Figure 3c.ii (estimated for both genders in a stacked regression). The final two columns show the results from main text Figure 3b.ii (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 Percentile: 5^{th}	Tide Percentile: 50^{th}	Tide Percentile: 95^{th}
0	-	-	-	-
	-	-	-	-
5	6e-04	-0.0048	0.0023	0.0089
	(0.00036)	(0.0046)	(0.0018)	(0.0037)
10	0.0013	-0.0081	0.0054	0.018
	(0.00053)	(0.0072)	(0.0027)	(0.0058)
15	0.0022	-0.0094	0.0092	0.026
	(0.00065)	(0.0095)	(0.0033)	(0.0076)
20	0.0033	-0.0087	0.014	0.035
	(0.00086)	(0.013)	(0.0044)	(0.0099)
25	0.0045	-0.0055	0.02	0.043
	(0.0012)	(0.017)	(0.0061)	(0.013)
30	0.006	0.00037	0.026	0.05
	(0.0017)	(0.02)	(0.0084)	(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 Figure 4. 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 4a. 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 4d. 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 systematically located in the low elevation areas within PIN codes, we compile a 100 m $\times$ 100 m grid cell level dataset recording each grid cell's average elevation (Methods), and the proportion of each cell which is classified as a slum location by Mumbai's governing civic body, the Brihanmumbai Municipal Corporation.[28]

In Table C.3, we present results from a grid cell level regression of elevation on a binary indicator for whether the grid cell is classified as a slum.

In columns (1) and (2), a basic regression shows that on average, slums are located in higher elevation areas of Mumbai. In column (2), we weight the regression using 2005 grid cell level population from the Global Human Settlement Layer[76] (Methods)[29]. However, this relationship disappears and becomes 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.

[28] We obtained a shapefile used by ref.[75] for the locations of slums in Mumbai through direct correspondence with the authors.

[29] We align the GHSL and elevation data onto a common 100m $\times$ 100m grid, by taking the average elevation in the intersecting grid cells. To calculate slum cover, we calculate the proportion of the grid cell classified as a slum according to our shapefile.

Table C.3: Elevation and slum location

Dependent Variable:	Elevation (m)			
Elevation (m)	(1)	(2)	(3)	(4)
Explanatory Variable:				
Slum Indicator	19.63***	17.40***	-4.133	0.7899
	(2.602)	(2.116)	(5.199)	(0.9322)
Fixed-effects				
PIN code	No	No	Yes	Yes
Regression weights	None	Population	None	Population
Observations	51,706	33,950	51,706	33,950

Standard-errors, which are clustered by PIN code, are reported in parentheses.
*Significance Codes: ***: p<0.01, **: p<0.05, *: p<0.1*

D Robustness checks

This appendix describes a series of robustness checks. Throughout, we reproduce the results from the main text in blue, alongside the robustness checks in grey. We show that our results are similar when estimated using different temporal aggregations of the mortality data (Section D.1), functional forms for the mortality-rainfall response (Section D.2), resolutions of the rainfall data (Section D.3), estimating samples (Section D.4), estimation strategies (Section D.5), control variables (Section D.6), location assignment of deaths (Section D.7), fixed effects (Section D.8), and weather data processing (Section D.9). Additionally, in Section D.10, we pay special attention to possible spillover effects of rainfall across time and space.

D.1 Mortality data at weekly vs. daily resolution

In our main model (Methods Equation 1), we estimate the mortality-rainfall response function using data that has been aggregated to the weekly level, with five weeks of rainfall (contemporaneous and 4 lagged weeks).[30] On one hand there are several advantages to estimating the model using weekly data, rather than through an analogous regression using daily data with 35 days of rainfall (contemporaneous day and 34 lagged days). First, it is more computationally efficient. Second, it is possible that our deaths data could have reporting biases, such that deaths are recorded on a different day than they actually occur. On the other hand, aggregating to the weekly level could introduce noise, since it allows for spurious correlations between future daily rainfall within a week and current mortality. Here we show that the estimated response function to daily rainfall is virtually identical when we estimate the regression at the daily vs weekly level (Figure D.1).[31]

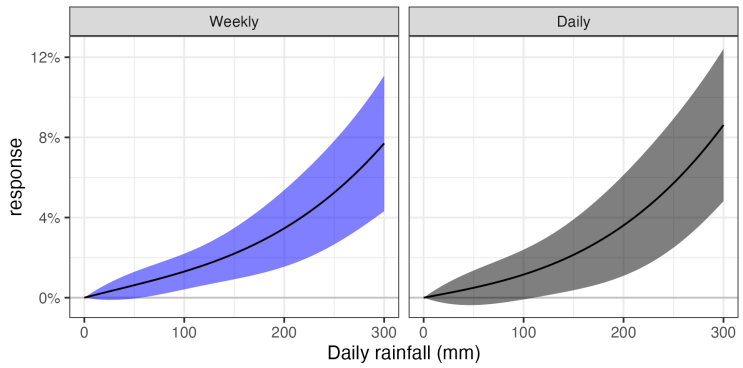


Figure D.1: Percent change in 5-week mortality due to daily rainfall: estimated at the daily vs weekly level.
Mortality-rainfall response functions are estimated for the full sample. The left hand side shows our main model, which is a weekly level regression with 4 weekly lags (main text Figure 2b). The right hand panel shows results from a model that is estimated at the daily level, and includes a 3rd order polynomial in contemporaneous daily rainfall, and 34 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.

[30] Note that since we take non-linear transformations of daily rainfall before aggregating to the weekly level, our response function traces out the impact of a day's rainfall[79].

[31] Regressions at the daily level include PIN code × year fixed effects and date (e.g., July 14, 2013) fixed effects.

D.2 Functional form of mortality-rainfall response

The results in main text Figures 2 and 3) estimate the mortality-rainfall response function using a third order polynomial in daily rainfall. Here we compare these results to those based on other functional forms—linear, quadratic polynomial, quartic polynomial, and a non-parametric function that bins daily rainfall into 50 mm wide intervals (Figure D.2). We find that the third order polynomial accurately yet parsimoniously captures the non-linearity seen in the binned function. This suggests that our choice of a functional form does not introduce an undue extrapolation of the underlying response function.

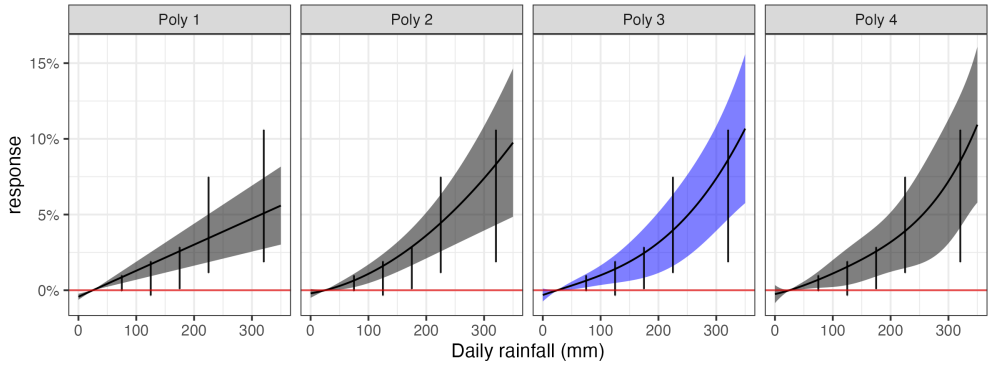


Figure D.2: Percent change in 5-week mortality due to daily rainfall: various polynomial and binned functional forms.

Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that use polynomials or bins 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[81]. 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 Resolution of rainfall data at which non-linear transformations are applied

The mortality-rainfall response function specified in Methods Equation (2) is a function of the daily rainfall totals in each PIN code. We choose the day as the fundamental unit of rainfall, since daily rainfall is salient for public and private decision making. Here we alternatively construct the mortality-rainfall response function using non-linear transformations of hourly or weekly rainfall totals (note, all regressions are still run after aggregating to the weekly level).[32] Results for hourly rainfall (Methods Equation 5) are shown in Figure 4a. We find results for weekly rainfall to be qualitatively similar to those estimated using the daily and hourly rainfall models. However, the weekly results are noisier, and less stable across the choice of linear, quadratic, cubic, and quartic response functional forms (Figure D.3).

[32] In the weekly case, the function would be constructed as follows, using the notation from Methods Equation 2: $\sum_{k=1}^{3}\sum_{l=0}^{4}\beta_{l,k}^{\text{weekly}}\left(\sum_{d\in w-l}R_{i,d,w-l,y}\right)^{k}$. The non-linear transformations (polynomials) are constructed on weekly rainfall totals (i.e., sums of daily rainfall totals over days in the week).

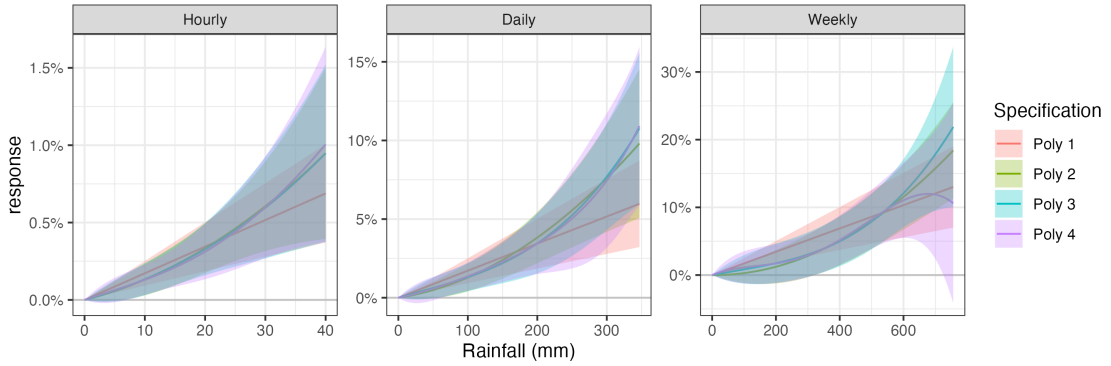


Figure D.3: Percent change in 5-week mortality due to daily rainfall: non-linearity taken at hourly vs daily vs weekly level.
Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses polynomials in hourly, daily, or weekly rainfall for periods in the 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.

D.4 Estimating sample

Here we examine the effect of varying the estimating sample. The rainfall data are available for a subset of the months in each year. The exact time span of the data varies within each year, with some years containing all 12 months and other years containing monsoon months only (see Table B.2). We show that our results are qualitatively robust to limiting the months used for estimation to those in and around the monsoon (Figure D.4). Thus our results are not driven by the unbalanced representation of months across years in the sample.

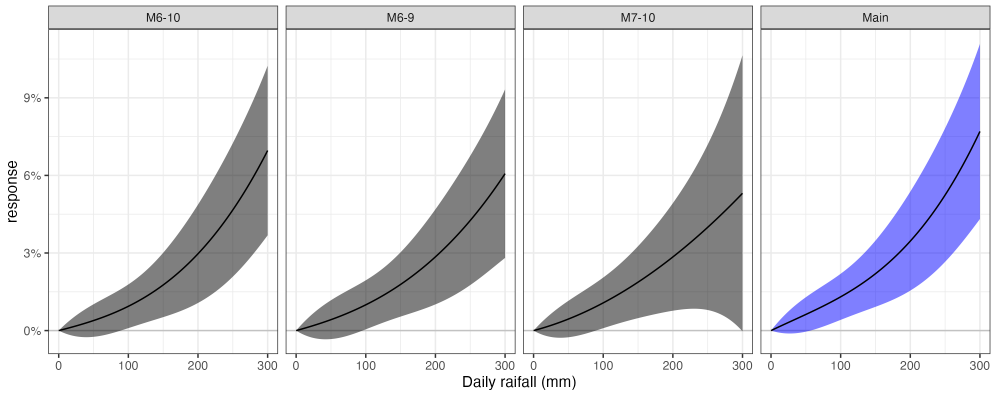


Figure D.4: Percent change in 5-week mortality due to daily rainfall: month of year sub-samples.
Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each curve represent the percent change in 5-week mortality due to a 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 main text Figure 2b. 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.5 Estimation strategy

Since our outcome variable, the number of weekly deaths, is a non-negative count variable, we use Poisson regression as our main estimation strategy[78]. However, here we show that our results are similar when using other estimation strategies (Figure D.5). These include ordinary least squares regression (OLS) with either weekly deaths or the natural log of weekly deaths as the outcome variable. In the latter case, we drop PIN-code $\times$ weeks that record zero deaths. We also show that our results are robust to population weighting the regression by PIN code population.

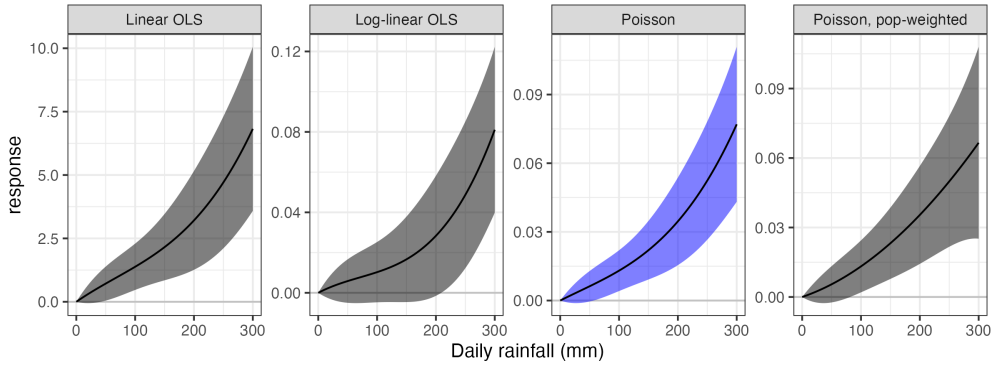


Figure D.5: Change in 5-week mortality due to daily rainfall: various estimation strategies.
Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along 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 an ordinary least squares model with the mortality count as the dependent variable; curve represents the change in mortality count. For the other panels, points along the curve represent the proportional change in mortality. Log-linear OLS indicates an ordinary least squares model with the natural log of the mortality count as the dependent variable, which requires dropping observations with zero deaths. The third facet shows results from the Poisson regression, reproducing main text Figure 2b. The final facet 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 Global Human Settlement Layer satellite data, see Methods). The curves in the three rightmost facets represent a percent change in mortality.

D.6 Temperature controls

Here we show that adding PIN code level temperature as a control variable in Methods Equation 1 does not alter our estimates.[33] This is not unexpected, since among the variables the date fixed effects in Equation (1) control for is city-level average temperatures in a given week. Because the 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, 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 (Figure D.6), and that the mortality effects of temperature are imprecisely estimated (Figure D.7).

[33]Specifically, we control for a third order polynomial of daily temperature aggregated to the weekly level, in the contemporaneous and 4 lagged weeks, analogous to what we do for rainfall in Methods Equation 2.

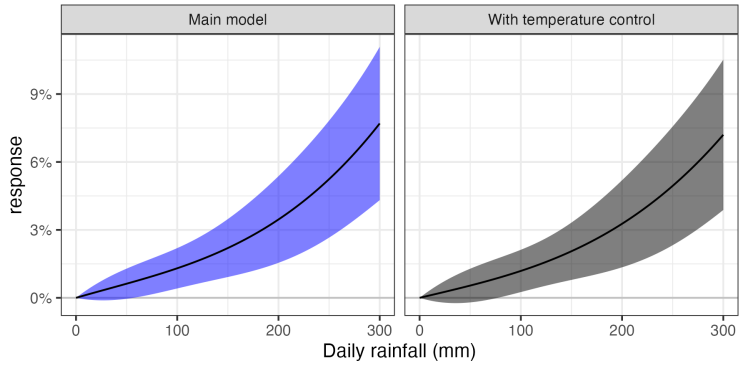


Figure D.6: Percent change in 5-week mortality due to daily rainfall: with and without temperature controls.

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 main text Figure 2b. 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.

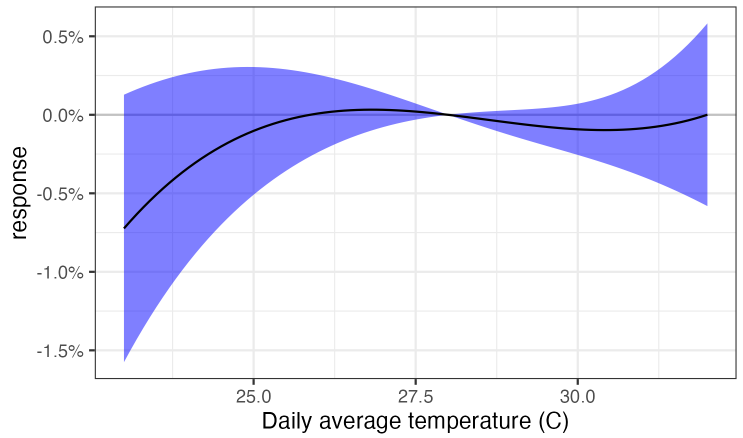


Figure D.7: Percent change in 5-week mortality due to daily temperature.

Mortality-temperature response function is estimated for the full sample. Estimates are from a model that uses a 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.6). 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.7 Location assignment

In our main model, we assign each death to the PIN code listed in the residential address of the deceased. 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. Here we show that our results

are robust to using different strategies to assign deaths to locations (Figure D.8).

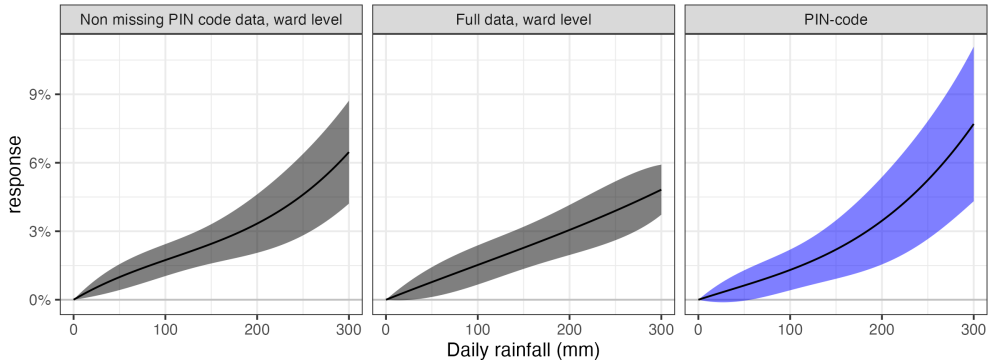


Figure D.8: Percent change in 5-week mortality due to daily rainfall: assigning location using different methods

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 main text Figure 2b, 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 lacking a Mumbai PIN code.

D.8 Choice of fixed effects

Here we show that our results are robust to using alternative fixed effects that are even more saturated than those in Methods Equation 1. We consider a specification in which we include date-by-area fixed effects in addition to the year-by-PIN code fixed effects in our main specification. The areas approximately correspond to each of Mumbai's main watersheds (Figure D.12), and the date-by-area fixed effects account for all area-level shocks. We also consider a specification that accounts for PIN-code specific seasonality (i.e, by including PIN code $\times$ week-of-year fixed effects), and one that interacts the PIN-code by year fixed effect with a monsoon season indicator, which accounts for PIN code-specific trends that vary for monsoon and non-monsoon months. Our results are robust across each of these specifications (Figure D.9).

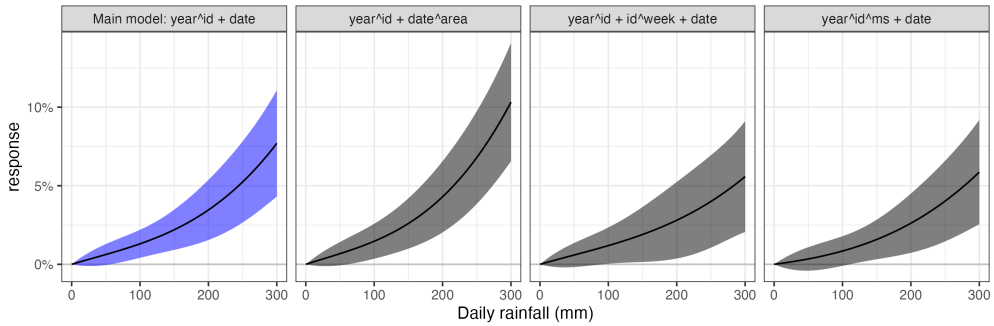


Figure D.9: Percent change in 5-week mortality due to daily rainfall: various fixed effects
Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along 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 main text Figure 2b, 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.12. 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.9 Alternative rainfall data processing steps

Here we conduct robustness checks where we make alternative decisions on the processing steps involved in converting our raw station level rainfall data to PIN code-by-day statistics. Specifically, we show our results are robust to using only the subset of stations that are available in every year of our sample. We also show our results are robust to not doing any interpolation or clipping (of extreme and implausible values) of the raw data (Appendix B.2.1). Thus, our results are not an artifact of the decisions we made in processing the AWS data.

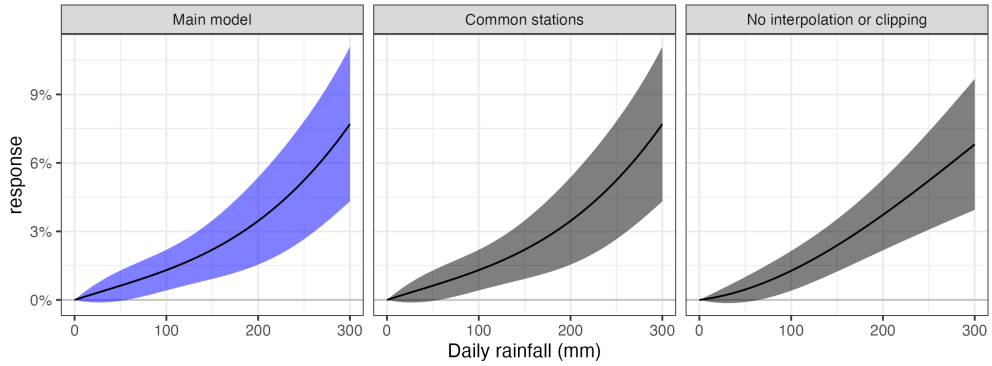


Figure D.10: Percent change in 5-week mortality due to daily rainfall: various AWS data processing procedures

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 leftmost panel reproduces main text Figure 2b, in which we process the AWS data as outlined in Appendix B.2. 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 rightmost 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.10 Spillover effects across time and space

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 spillover effects across time and space. Here we investigate these possibilities, and conclude that our main model likely understates the true causal effect due to long lasting spillover effects over time (Section D.10.1). We do not find meaningful evidence of spillovers across space (Section D.10.2).

D.10.1 Temporal spillovers

Our main model 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 2b 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 main text Figure 2c, we present the breakdown of the effect for each of the 5 weeks and see that it is roughly similar in each week.[34] This suggests that while there are immediate mortality effects of 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 does not simply advance the timing of deaths without altering the total number of deaths.

[34] The height of the curve in main text Figure 2b is equal to the average of the height of the 5 curves in main text Figure 2c. The averaging is necessary to normalise percent increases in each week's mortality into a percent increase in mortality over a 5-week period.

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

Figure D.11 shows the results from estimating regressions of daily mortality on daily rainfall, varying the number of lags of daily rainfall variables included.[35] The left hand-side panel shows the sum over all lagged response functions 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 at 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.

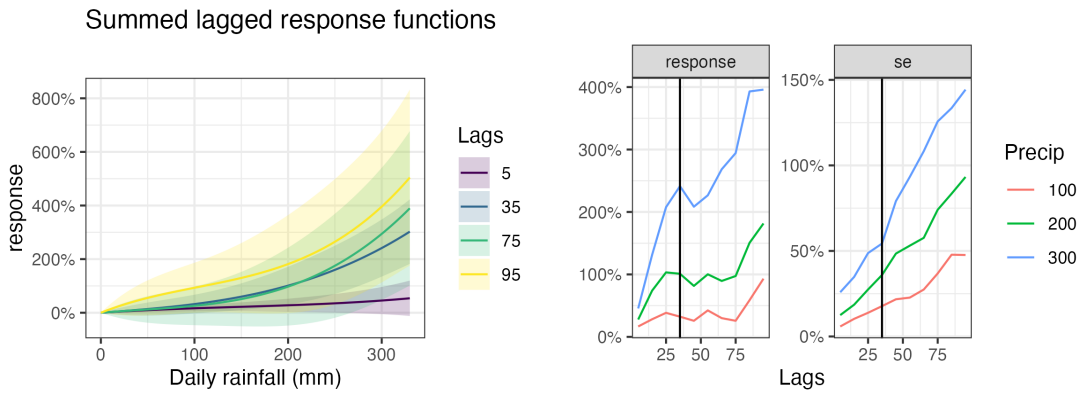


Figure D.11: Percent change in mortality due to daily rainfall: daily regressions with increasing numbers of lags.

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 figures, 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 specific daily rainfall values– 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 rainfall, then we should expect the sum of the lagged response functions to eventually start tending down towards zero, as the only effect of the 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.11 suggests that a specification that accounts for roughly 35 days of lagged rainfall understates 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 (Appendix D.1).

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

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 analysis, 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.10.2 Spatial spillovers

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

The randomisation check in main text Figure 1d suggests that spillover effects across space are unlikely to be large. In particular, we fail to find an effect of rainfall on mortality when estimating the model with rainfall values that have been randomly shuffled across PIN codes on the same date (main text Figure 1d, middle plot). If the effects of local rainfall are propagated throughout the city, we would expect to find a positive effect even when estimating with these shuffled values.

Nevertheless, we also implement two additional strategies to check for possible spatial spillovers. First, we estimate the response function at various levels of spatial aggregation. Second, we estimate a model in which we separately control for rainfall in neighbouring PIN codes.

Estimates at different levels of spatial aggregation

Here we estimate the response function based on Methods Equation 1, but using mortality and rainfall data at more highly aggregated spatial units than PIN codes. These coarser spatial units— "wards" and "areas"— consist of groupings of PIN codes and are illustrated in the maps in Figure D.12.[36] Estimation using coarser spatial units captures the net spillover effects across PIN codes within those units. As in our main specification, we use a third-order polynomial in daily rainfall for days in the contemporaneous week and 4 lagged weeks. For an aggregated spatial unit j (ward or area) consisting of N_j PIN codes, the rainfall response function f is constructed as follows:

$$f(\boldsymbol{R}^{\text{daily}}_{j,w,y}) = \sum_{k=1}^{3}\sum_{l=0}^{4} \beta^{\text{daily}}_{l,k} \sum_{d \in w-l} \left(\frac{1}{N_j} \sum_{i \in j} R^{k}_{i,d,w-l,y} \right), \tag{D.1}$$

where total daily rainfall values on each day d, raised to the kth power, are averaged across the PIN codes in j.[37]

The curves in Figure D.12 depict rainfall response functions estimated at various spatial resolutions. For instance, at the PIN code level, a single day with 150 mm rainfall causes a 2.2% ($\pm$0.65%, p<0.001) increase in 5-week mortality in that PIN code (blue curve). At the ward level, a single day with 150 mm average rainfall

[36] For this purpose, we group PIN codes into "wards" based on the municipal ward with which they share the most overlap. We group PIN codes into "areas" by manual inspection of maps of Mumbai's watersheds. This grouping is intended to capture sets of PIN codes that are likely to experience spillovers due to run-off from each other's rainfall.

[37] As in Equation 2, each parameter, $\beta^{\text{daily}}_{l,k}$, denotes a coefficient on the k^{th} power and l^{th} lag term.

across PIN codes in the ward causes a 2.7% ($\pm$0.72%, p<0.001) increase in 5-week mortality in that ward. The similar magnitudes of these effects suggest that on net, there are minimal cross-PIN code spillovers in the mortality effects of rainfall. This finding is generally robust even at higher levels of spatial aggregation. At the area level, a single day with 150 mm average rainfall across PIN codes in the area causes a 2.3% ($\pm$1.1%, p=0.035) increase in 5-week mortality in that area.

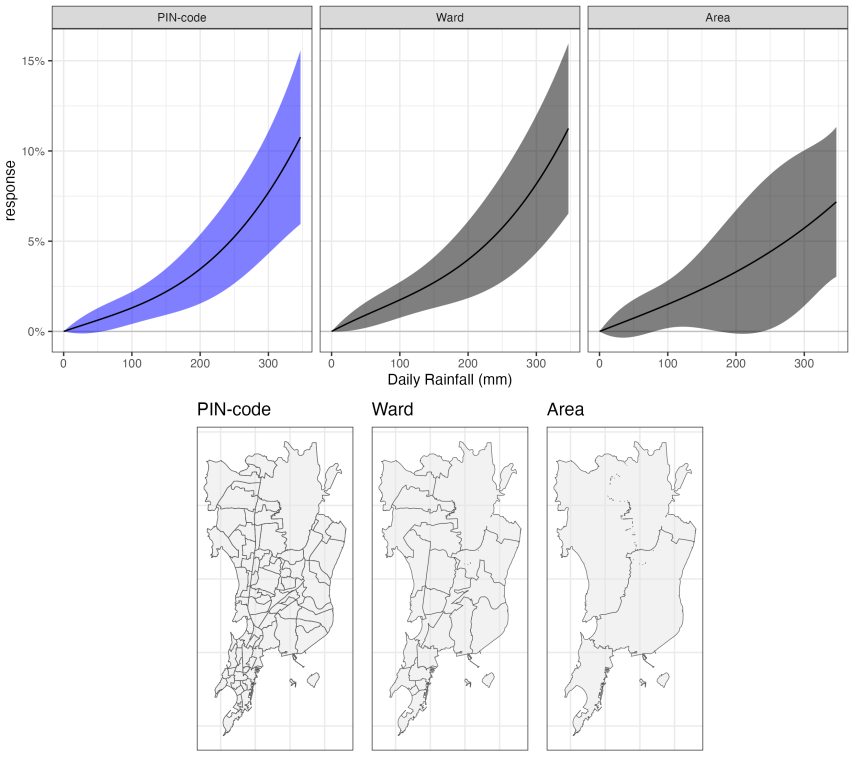


Figure D.12: Percent change in 5-week mortality due to daily rainfall: estimated at various levels of spatial aggregation.

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 main text Figure 2b. 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.

Thus, the response curves in Figure D.12 do not provide strong evidence of spatial spillovers across PIN codes within the same ward or area, while the estimates using rainfall values shuffled across PIN codes within date (main text Figure 1d, middle plot) suggest that city-wide spillover effects are unlikely to be large.

Controlling for rainfall in neighbouring PIN codes

We also investigate the possibility of spillover effects across space by estimating a version of Equation (1)

that includes a control for the mean daily rainfall levels in neighbouring PIN codes.[38]

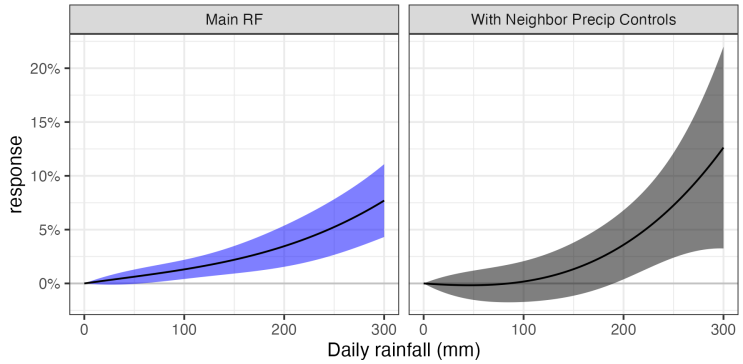


Figure D.13: Percent change in 5-week mortality due to daily rainfall: controlling for neighbouring rainfall

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 main text Figure 2b. The right panel shows a model that additionally contains a third-order polynomial in the mean daily rainfall across neighbouring PIN codes, 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 rainfall 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.13). Figure D.14 compares the mortality effects of own-PIN code and neighbouring PIN code rainfall. The effects of neighbouring PIN code rainfall are very imprecisely estimated.

[38]The mean daily rainfall in neighbouring PIN codes is raised to each polynomial order before being aggregated across days in the week, in the same manner as described in Methods Equation 2. We note that the specification uses the same fixed effects as in Methods Equation 1, including date fixed effects, which control for the city-wide average level of rainfall on each date.

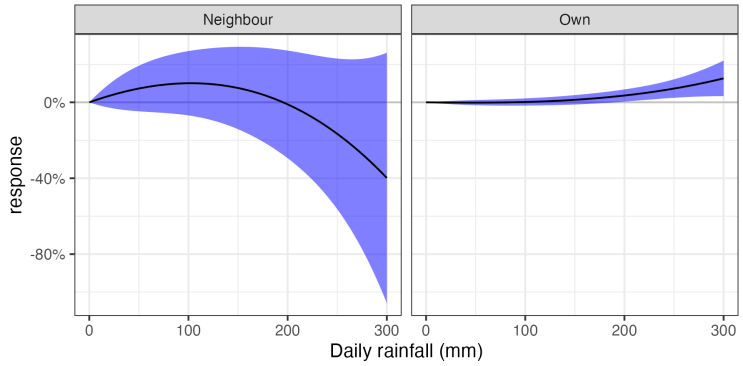


Figure D.14: Percent change in 5-week mortality due to daily rainfall: own vs. neighbouring PIN code rainfall

Mortality-rainfall response functions are estimated for the full sample, using a model that contains a third-order polynomial in own daily rainfall in the contemporaneous week and four lagged weeks, as well as the same terms for the mean daily rainfall across neighbouring PIN codes. 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 codes' 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.13, right panel. Shaded areas indicate 95% confidence intervals.

E Differential effects of rainfall on mortality by age, gender, residence, elevation, and tide height during rainfall

This appendix considers alternative ways to model differential effects of rainfall on mortality, including different definitions of age categories (Section E.1), intersections of age groups, genders, and slum/non-slum residence (Section E.2), and different measures of elevation (Section E.3) and tide (Section E.4).

E.1 Age categories

When estimating the mortality effect of rainfall by age group in main text Figure 2a.ii, we define the youngest age category as consisting of children under the age of 5. Here we consider alternative definitions of the youngest age category by varying its upper age limit from 1 to 20 in one-year increments (Figure E.1). We find a stronger mortality-rainfall relationship when we restrict the category to only include younger people. However, the response is similar in magnitude for all definitions of the youngest age categories that have an upper limit of up to 5 years. (Figure E.1, right plot).

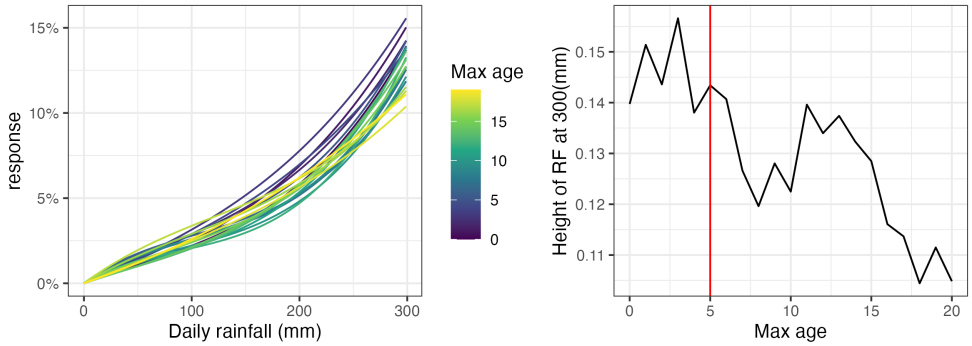


Figure E.1: Percent change in 5-week mortality due to daily rainfall: robustness under different categorisations of the minimum age group

Mortality-rainfall response functions are estimated for the full sample, using Methods Equation (3) with a differential response by age category. 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 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, variously defined as ages $[0, X]$, $X \in [0, 20]$. The right hand plot shows the height of the response function at 300 mm as a function of the maximum age in the youngest age category. The red line at 5 years denotes our main specification (main text Figure 2a.ii), where the youngest age category consists of persons under 5 years of age.

E.2 Differential effects for intersections of age groups, genders, and residence

Here we report estimates of the mortality-rainfall response differentiated by age $\times$ residence groupings (Section E.2.1), and for gender $\times$ residence groupings and age $\times$ gender $\times$ residence groupings (Section E.2.2). Estimates for these more granular subgroups tend to be imprecise.

E.2.1 Age × residence groupings

Main text Figure 3b.ii reveals that slum residents suffer higher increases in mortality due to rainfall than do non-slum residences. Age-wise mortality-rainfall responses for slum and non-slum residents (Figure E.2) reveal that this difference is primarily due to 5-64 and 65+ slum residents being harmed more than non-slum residents in those same respective age groups. The response for the under-5 age group is relatively similar for slum and non-slum residents, though is imprecisely estimated. It should be noted that overall in-sample deaths of children under-5 occur disproportionately among slum residents (Table E.1).

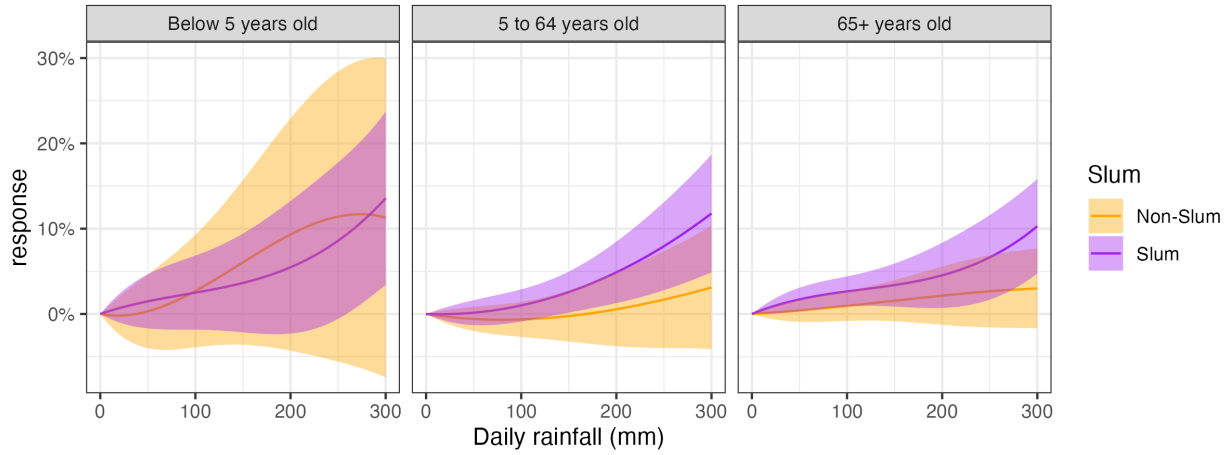


Figure E.2: Percent change in 5-week mortality due to daily rainfall: differential effect by age and slum residence.
Mortality-rainfall response functions are estimated for the full sample, using Methods Equation (3) with a differential age-wise response for slum and non-slum residents. Estimates are from a model that uses a third-order polynomial in daily rainfall in the contemporaneous week and four lags. Points along each 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.

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.

E.2.2 Residence, gender, and age

Gender-wise mortality-rainfall response estimates for slum and non-slum residents (Figure E.3) reveal that the female disadvantage seen in main text Figure 3c.ii is driven entirely by the non-slum population. Moreover, within the non-slum population, the female disadvantage is found most starkly among the 65+ age

group (Figure E.4).[39] The age composition of 65+ male and female deaths among slum and non-slum residents offers a possible explanation for this pattern. Density plots for age of death (Figure E.5) show that non-slum resident females tend to die at considerably older ages than non-slum resident males, particularly conditional on dying at the age of 65 or later. Because they tend to be older, females who die at 65 or later may have worse baseline health than males who died at 65 or later, making them more vulnerable to the harms brought by rainfall and urban flooding.

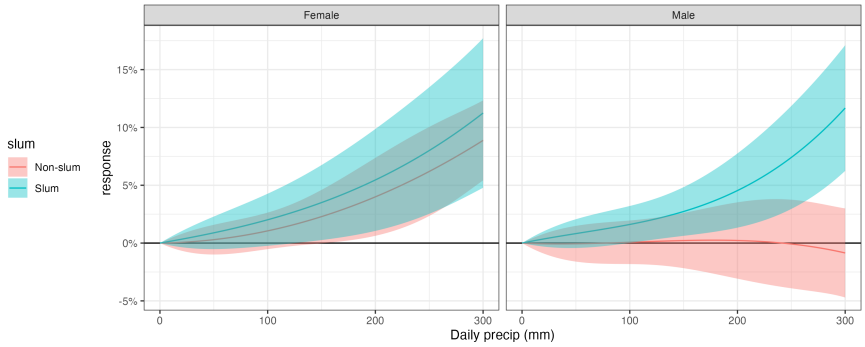


Figure E.3: Percent change in 5-week mortality due to daily rainfall: differential effect by gender and slum residence.

Mortality-rainfall response functions are estimated for the full sample, using Equation (3) with a differential gender-wise response for slum and 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 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.

[39] This suggests that maternal mortality during or following childbirth cannot explain the male-female differential in mortality response to rainfall.

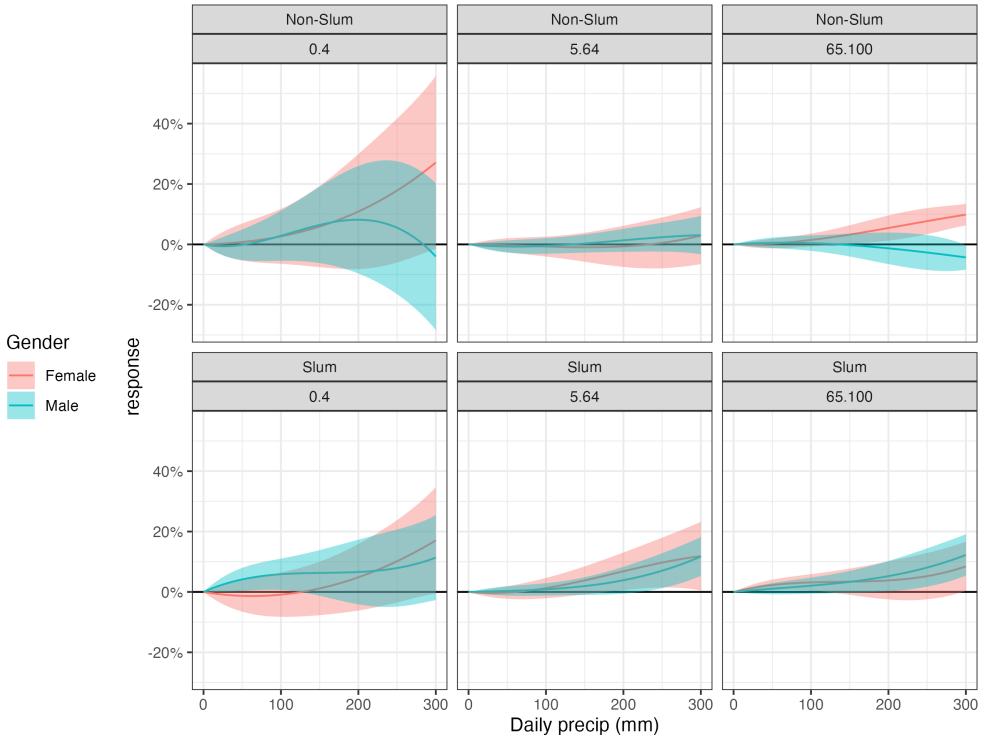


Figure E.4: Percent change in 5-week mortality due to daily rainfall: differential effect by gender, age and residence.

Mortality-rainfall response functions are estimated for the full sample, using Equation (3) with a differential response by gender, age, and residence. 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.

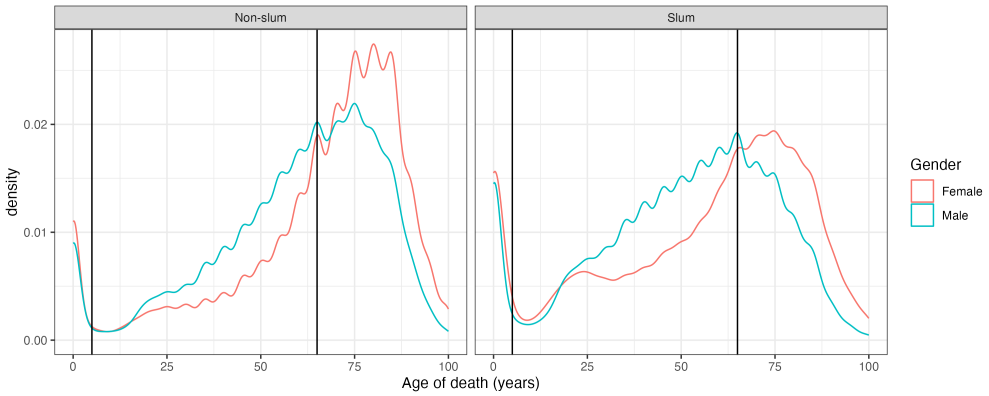


Figure E.5: Age distribution of deaths by slum residence and gender.

Curves plot the kernel density of age of death for male and female slum residents and non-slum residents. Vertical lines indicate 5 years and 65 years.

E.3 Measures of elevation

Given the importance of drainage patterns and topology in determining urban flooding, we expect that lower elevation PIN codes are more at risk of harms from rainfall. To test this, we split the PIN codes in our sample into two groups according to either their mean elevation level (as done in main text Figure 3b.iii), mean height above nearest drainage (HAND)[40], or percent of land area below 5 metres elevation. Figure E.6b shows that under all three definitions, the mortality increase due to rainfall is higher in low elevation PIN codes than in high elevation PIN codes. This difference appears to be driven mainly by the 5-64 and 65+ age groups (Figure E.6c).

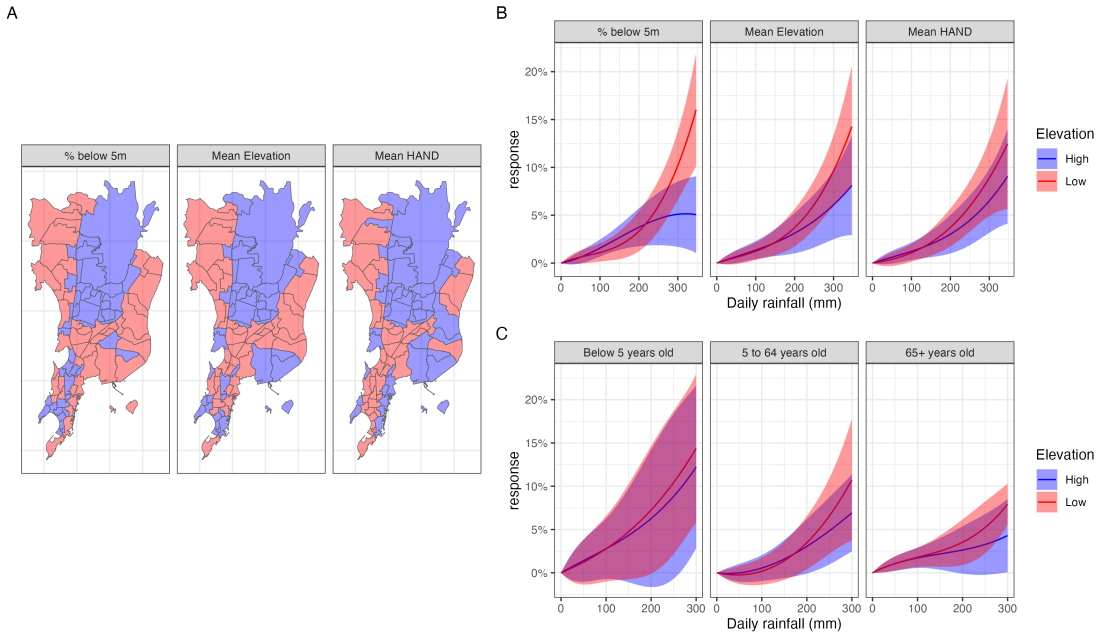


Figure E.6: Percent change in 5-week mortality due to daily rainfall: low vs. high elevation PIN codes.

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 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. 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 5 m above sea level. 'Mean Elevation' 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 mortality-rainfall responses for high and low elevation PIN codes according to each of these measures, with the plot based on mean elevation reproducing the dashed curves in main text Figure 3b.ii. Finally, Panel (C) shows age-wise mortality-rainfall responses for high and low elevation PIN codes as defined by mean elevation.

[40] "Height Above Nearest Drainage" (HAND) is calculated by ref.[73] and obtained through direct correspondence.

E.4 Measures of tide

In main text Figure 4d we show that hourly rainfall causes higher mortality when it occurs during a high predicted tide (Methods Equation 6). Here we consider alternate ways to model the coincidence of rainfall and high predicted tides. In Section E.4.1 we examine the impact of rainfall occurring during a predicted tide level above or below a threshold of 450 cm. In Section E.4.2, we examine the impact of rainfall by the average daily predicted tide height on the day it occurs.

E.4.1 Tide threshold

Let $\tau_{h,d,w,y}$ denote the hourly predicted 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 predicted 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. Figure E.7 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 hourly response function, presented in main text Figure 4a, is included for reference in green in both panels.

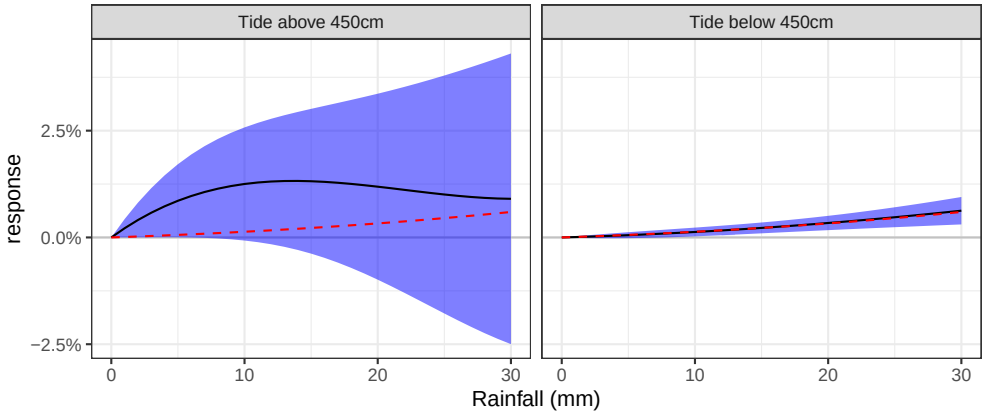


Figure E.7: Percent change in 5-week mortality due to hourly rainfall occurring during low and high tide periods.

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 main text Figure 4a.

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 tide $>$ 450 cm can quickly overwhelm Mumbai's drainage system.[41]

E.4.2 Daily average tide

Figure E.8 depicts an illustrative example of variation in predicted average daily tides over a month-long time span.

It is possible that flooding can occur when heavy rain falls on a day with high daily average tide height.[42] To investigate this, we estimate a model where we interact daily total rainfall with a measure of the predicted average daily tide value. Let $T_{d,w,y}$ denote the daily average tide on day d, week w and year y, after normalising it to be mean zero and have standard deviation 1. We augment our main response function (Methods

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

[42] For example, the Mumbai municipal government 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 E.8 depicts tidal patterns in June 2008, illustrating fluctuations in tidal heights and amplitudes.

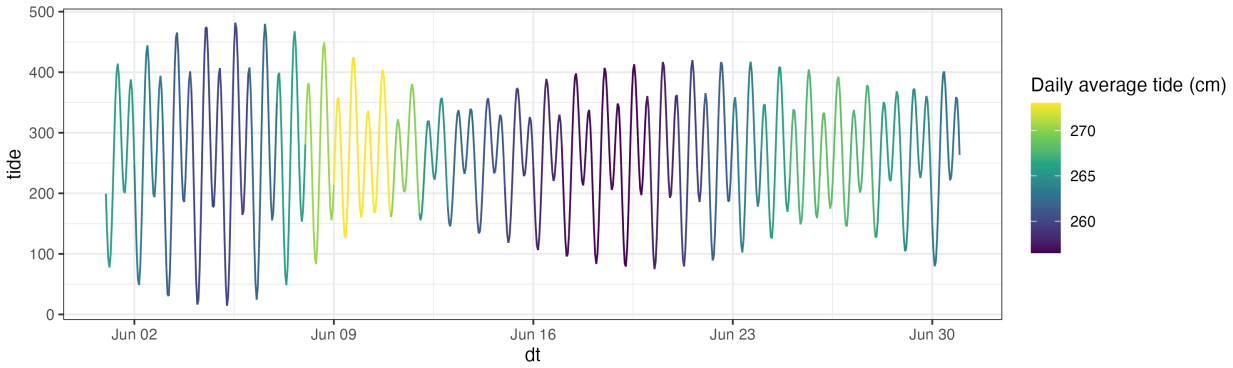


Figure E.8: Daily average predicted tide levels during June 2008.
Graph plots time series of predicted hourly tide heights over a single month (June 2008). Colouring indicates the daily average value over the hours for each day.

Equation (2)) as follows:

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

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

Figure E.9 presents the results of estimating this model. We find that when the predicted 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.

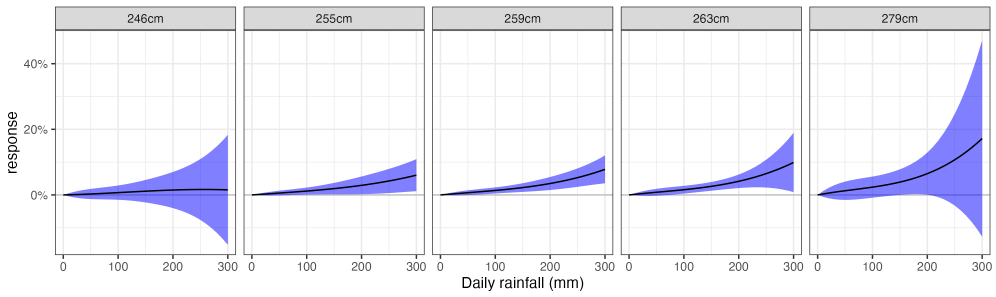


Figure E.9: Percent change in 5-week mortality due to daily rainfall: linear daily tide interaction.
Mortality-rainfall response functions are estimated for the full sample. Estimates are from a model that uses a third-order polynomial in daily rainfall and a interactions of these polynomial terms with normalised predicted daily average tide height, 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. Each facet shows the estimated response function evaluated at different levels of the normalised daily tide value (from left to right: 2 standard deviations below mean, 1 standard deviation below mean, mean, 1 standard deviation above mean, 2 standard deviations above mean). The tide heights in centimetres associated with each of these normalised values are displayed at the top of each facet.

F The mortality burden of rainfall: Alternative models and confidence interval calculation

In this appendix, we present additional results for the mortality burden of rainfall based on models with alternative fixed effects estimated in Section D.8 (Section F.1). In addition, we detail the procedure used to calculation confidence intervals for the mortality burden of rainfall (Section F.2).

F.1 Alternative models

Table F.1 reproduces the results on the mortality burden of rainfall that are presented in main text Figure 5a, for the population-wide model as well as age-wise, residence-wise, and gender-wise models. In addition, in Table F.2 we present estimates of the mortality burden of rainfall using models with alternative fixed effects (Section D.8). Across all models, we find a similar magnitude of mortality burden.

	Age Category			**Residence**		**Gender**	
Total	*<5*	*5-64*	*65+*	*Non-slum*	*Slum*	*Male*	*Female*
Population-wide model							
0.084 (0.01, 0.17)							
Age-wise model							
0.074 (0.0052, 0.16)	0.17 (-0.083, 0.57)	0.0086 (-0.1, 0.14)	0.12 (0.03, 0.23)				
Residence-wise model							
0.071 (-0.0063, 0.16)				0.02 (-0.094, 0.15)	0.11 (0.0069, 0.23)		
Gender-wise model							
0.083 (0.014, 0.18)						0.059 (-0.056, 0.2)	0.11 (0.0071, 0.24)

Table F.1: Proportion of monsoon season deaths caused by rainfall.
Table shows the excess mortality from rainfall as a proportion of overall mortality for the full population and by age, residence and gender during the monsoon season, reproducing results from main text Figure 5a. The first row displays results based on the population-wide model estimated estimated in main Figure 2b. The second, third, and fourth rows respective display results based on the age-wise, residence-wise, and gender-wise model, respectively estimated in main text Figures 3a.ii, 3b.ii, 3c.ii. Point estimates are calculated based on Equation (7); 95% confidence intervals, calculated with a parametric bootstrap described in Appendix F.2, are in parentheses.

Fixed effects	Mortality burden
PIN code×year + week×year	0.084 (0.01, 0.17)
PIN code×year + week×year×area	0.082 (0.0027, 0.18)
PIN code×year×season + week×year	0.05 (-0.039, 0.15)
PIN code×year + PIN code×week + week×year	0.078 (-0.012, 0.17)

Table F.2: Proportion of monsoon season deaths caused by rainfall: models with alternative fixed effects.

Table shows the excess mortality from rainfall as a proportion of overall mortality for the full population during the monsoon season. The first row reproduces the result from our main model, which uses PIN code × year and week × year (i.e., date) fixed effects (first row in Table F.1). The second row shows the result based on a model in which the week × year fixed effects are interacted with an indicator representing three areas of Mumbai, which are the same as those used in Figure D.12. The third row shows the result based on a model in which the PIN code × year fixed effects are interacted with an indicator for whether the month-of-year is between June and September (inclusive, i.e. monsoon vs. non-monsoon season), and also uses week × year fixed effects. The fourth row shows the result based on a model with PIN code × year, PIN code × week-of-year, and week × year fixed effects. Point estimates are calculated based on Methods Equation (7); 95% confidence intervals, calculated with a parametric bootstrap described in Appendix F.2, are in parentheses.

F.2 Confidence intervals for the mortality burden of rainfall

The procedure for characterizing uncertainty in our estimates of the mortality burden of rainfall is as follows:

1. Estimate the mortality-rainfall response model, save the coefficients and variance-covariance matrix. Calculate the point estimate for r using Methods Equation 7.
2. Take 1000 bootstrap draws from the joint normal sampling distribution of the coefficient estimates.
3. Calculate $r_{i,w,y,k}$ for each bootstrap draw k for each PIN code×week×year in our sample.
4. For each bootstrap draw k, sum these estimates across PIN codes and weeks in our sample.
5. Calculate quantiles of the distribution (across bootstrap draws) of the sum calculated in Step 4 to obtain for the upper and lower bounds of confidence interval (i.e., 97.5^{th} and 2.5^{th} percentiles in the case of a 95% confidence interval).
6. Divide upper and lower confidence interval bounds by the total deaths observed in the sample in order to express the values as proportions.