Title: Comparing lagged impacts of mobility changes and environmental factors on COVID-19 waves in rural and urban India: a Bayesian spatiotemporal modelling study

Authors: Eimear Cleary[1*†], Fatumah Atuhaire[1], Alessandro Sorcihetta[2], Nick Ruktanonchai[3], Cori Ruktanonchai[3], Alexander Cunningham[1], Massimiliano Pasqui[4], Marcello Schiavina[5], Michele Melchiorri[5], Maksym Bondarenko[1], Harry E R Shepherd[1], Sabu S Padmadas[6,7], Amy Wesolowski[8], Derek A T Cummings[8,9], Andrew J Tatem[1†], Shengjie Lai[1,10*†]

1.WorldPop, School of Geography and Environmental Science, University of Southampton, UK

2. Department of Earth Sciences "Ardito Desio", Università degli Studi di Milano, Milan, Italy

3. Department of Population Health Sciences, VA-MD College of Veterinary Medicine, Virginia Tech, USA

4. Institute for Bioeconomy, National Research Council of Italy (IBE-CNR), Rome, Italy

5. European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra, VA, Italy

6. Department of Social Statistics & Demography, Faculty of Social Sciences, University of Southampton, UK

7. Department of Public Health & Mortality Studies, International Institute for Population Sciences, Mumbai, India

8. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

9. Department of Biology and Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

10. Institute for Life Sciences, University of Southampton, Southampton, UK

*These authors contributed equally

†Corresponding author. Email: Shengjie.Lai@soton.ac.uk (S.L.); e.cleary@soton.ac.uk (E.C.) A.J.Tatem@soton.ac.uk (A.J.T.)

Abstract

Previous research in India has identified urbanisation, human mobility and population demographics as key variables associated with higher district level COVID-19 incidence. However, the spatiotemporal dynamics of mobility patterns in rural and urban areas in India, in conjunction with other drivers of COVID-19 transmission, have not been fully investigated. We explored travel networks within India during two pandemic waves using aggregated and anonymized weekly human movement datasets obtained from Google, and quantified changes in mobility before and during the pandemic compared with the mean baseline mobility for the 8-week time period at the beginning of 2020. We fit Bayesian spatiotemporal hierarchical models coupled with distributed lag non-linear models (DLNM) within the integrated nested Laplace approximate (INLA) package in R to examine the lag-response associations of drivers of COVID-19 transmission in urban, suburban, and rural districts in India during two pandemic waves in 2020-2021. Model results demonstrate that recovery of mobility to 99% that of pre-pandemic levels was associated with an increase in relative risk of COVID-19 transmission during the Delta wave of transmission. This increased mobility, coupled with reduced stringency in public intervention policy and the emergence of the Delta variant, were the main contributors to the high COVID-19 transmission peak in India in April 2021. During both pandemic waves in India, reduction in human mobility, higher stringency of interventions, and climate factors (temperature and precipitation) had 2-week lag-response impacts on the R_t of COVID-19 transmission, with variations in drivers of COVID-19 transmission observed across urban, rural and suburban areas. With the increased likelihood of emergent novel infections and disease outbreaks under a changing global climate, providing a framework for understanding the lagged impact of spatiotemporal drivers of infection transmission will be crucial for informing interventions.

Keywords: COVID-19; human mobility; non-pharmaceutical interventions; travel restriction; climate; INLA; DLNM; Bayesian spatiotemporal model; lag-response; India.

Introduction

The COVID-19 pandemic highlighted the intrinsic role of human movement, along with demographics and environmental factors, in the dispersal of human pathogens in a highly connected, mobile and globalised society.[1-3] As the global climate changes and environmental and extreme weather events increase in frequency, emergence of novel zoonotic diseases and outbreaks of bacterial, parasitic and viral infections are likely to become more frequent[4]. Effective and efficient responses to future outbreaks and epidemics require a thorough understanding of the infection transmission drivers that contributed to different COVID-19 pandemic waves. Furthermore, it is essential to develop a framework for examining the spatiotemporal variations in transmission drivers across urban, suburban, and rural areas.

In India, the initial wave of COVID-19 was contained by a nationwide lockdown, which extended from March 31st to May 31st, 2020,[5] with a subsequent phased lockdown for containment zones in effect until June 30th, 2020.[6] The first wave of COVID-19 transmission in India was characterised by mild clinical infection and a relatively low mortality rate of less than 3%.[5] Several serosurveys carried out following the initial pandemic wave in India determined a high proportion of asymptomatic infections[7-10], leading to speculation as to the reasons for lower incidence of severe clinical cases including population demographics and innate population immunity.[11,12]

In March 2021, India experienced a severe second wave of COVID-19 transmission with a high proportion of infection associated mortality.[13] The Delta variant, or B.1.617 lineage, dominant during the second transmission wave was first identified in Maharashtra in late 2020[14] before quickly spreading throughout India and to at least 90 other countries.[15]

Compared with the initial pandemic wave in India, the Delta wave was characterised by high morbidity and mortality, even among a younger age cohort, overwhelming health systems across the country.[16,17] On April 26th 2021, India recorded 360,960 new cases, at the time the highest number of daily new SARS-CoV-2 infections recorded worldwide,[18] and by mid-June 2021 more than 29 million cases of COVID-19 had been confirmed.[19] During the second pandemic wave, the number of COVID-related deaths in India ranked third globally with an estimated 2.7 million COVID-19-related deaths occurring between April and July 2021.[20]

Although reasons for the second wave of transmission were unclear, it was speculated that the surge in case numbers was attributed to the circulation of the B.1.617 lineage of SARS-CoV-2 (Delta variant), which had a more effective transmission capability, shorter incubation period and was more pathogenic than previous lineages.[17,21,22] Prior to this surge in transmission, adherence to COVID-19 preventative behaviours in India was less stringent, likely due to pandemic fatigue, economic necessity and complacency due to the perception that clinical case infections in India were mild relative to other populations.[16,23] Population mobility, which had begun to increase relative to mobility during national lockdown interventions, including rural-urban-rural migration to mass election rallies and social and religious gatherings such as Kumbh Mela (approximately 7 million people), was also likely to be a primary driver of the second wave of SARS-CoV-2 in India.[13,15]

Previous research has explored the relationship between human mobility in response to government interventions and COVID-19 transmission during the early stages of the pandemic [24,25,26], or state level associations between human mobility and COVID-19 transmission during the Delta pandemic wave.[27] However, no previous research has compared mobility patterns, or inter-district movement across both pandemic waves, relative to pre-pandemic mobility levels, and associated impact on COVID-19 transmission. The contribution of district level urbanisation,[28,29] population density and demographics,[30,31] climate[32,33] and stringency of

government interventions[24] to COVID-19 transmission in India has also previously been investigated. However methodological approaches have included simple correlation[24,30] or regression analyses[34] and, to the best of our knowledge, no spatiotemporal modelling approach has been used to explore the urban-rural district level associations of human mobility, stringency of government intervention, and climate to transmission risk across both pandemic waves in India.

In this study, we quantified changes in mobility patterns and travel networks across India, before and during the COVID-19 pandemic, using spatially resolved, aggregated and anonymized weekly human movement datasets obtained from Google. We used a Bayesian spatiotemporal hierarchical framework, coupled with distributed lag non-linear models (DLNM) to examine the lag-response associations between the transmission dynamics of COVID-19 and drivers of transmission during the initial wave (July to November 2020) and Delta wave (March to July 2021) of SARS-CoV-2 in India. We also compared the lagged impacts of mobility metrics, climate covariates, and stringency of government interventions on the transmission of SARS-CoV-2 lineages between both pandemic waves, and across urban, suburban and rural delineated districts.

Methods

Data sources

Covid-19 incidence data

In India, administrative units are divided in state (36 including eight union territories), district and township, corresponding to spatial administrative levels I, II and III, respectively (**SI Figure S1**). The daily number of confirmed COVID-19 cases at country level were obtained from the Data Repository assembled by the Centre for Systems Science and Engineering (CSSE) at Johns Hopkins University[35]. We also obtained COVID-19 data at district (admin II) level for the period from 26 April 2020 to 31 October 2021 for 666 districts from

www.covid19india.org, a volunteer driven, crowdsourced tracker for COVID-19 cases in India.[36] COVID-19 data were available in 666 district units, as in some cases, depending on testing capacity and guidelines in each federal state, data were aggregated to state level only or case incidence estimated by state pool.[36]

Administrative level I and II shapefiles for India, corresponding with state and district level, were obtained from the Database of Global Administrative Areas (GADM version 3.6) (https://gadm.org/). Since the last national census of population in India in 2011, new districts have been created by splitting and rearranging some administrative boundaries.[30] COVID-19 data aggregated to current district boundaries were merged with 2011 administrative level II units according to the best spatial alignment of current and previous district boundaries. For the purpose of spatial modelling, the islands in Lakshadweep and the Andaman Islands have been unified as discrete spatial areas and treated as distinct districts. The authors remain neutral with regard to jurisdictional claims in maps used in this study.

Google COVID-19 Aggregated Mobility Research Dataset

Aggregated and anonymized weekly human movement datasets were obtained from Google to measure the changes of mobility across and within regions from November 10, 2019, to December 31, 2021, and to assess their impacts on COVID-19 transmission in India. The Google mobility dataset contains anonymized mobility flows aggregated over users who have turned on the Location History setting, which is off by default. This is similar to the data used to show how busy certain types of places are in Google Maps — helping to identify when a local business tends to be the most crowded. The dataset aggregates flows of people between *S2* cells, which here is further aggregated by district of origin and destination. Each S2 cell represents a quadrilateral on the surface of the planet and allows for efficient indexing of geographical data.

To produce this dataset, machine learning was applied to log data to automatically segment data into semantic trips.[37,38] To provide strong privacy guarantees, all trips were anonymized and aggregated using a differentially private mechanism to aggregate flows over time (see https://policies.google.com/technologies/anonymization). This research is done on the resulting heavily aggregated and differentially private data. No individual user data was ever manually inspected, only heavily aggregated flows of large populations were handled. All anonymized trips are processed at aggregate level to extract their origin, destination, location and time. For example, if users travelled from location a to location b within time interval t, the corresponding cell (a, b, t) in the tensor would be n∓err, where err is Laplacian noise. The automated Laplace mechanism adds random noise drawn from a zero mean Laplace distribution and yields (ϵ, δ)-differential privacy guarantee of $\epsilon = 0.66$ and $\delta = 2.1 \times 10{-}29$. The parameter ϵ controls the noise intensity in terms of its variance, while δ represents the deviation from pure ϵ-privacy. The closer they are to zero, the stronger the privacy guarantees. Each user contributes at most one increment to each partition. If they go from a region A to another region B multiple times in the same week, they only contribute once to the aggregation count.

The summed weekly domestic mobility inflows and outflows of each district were then divided by the number of origin S2 cells (each was calculated only once) that contained data between November 10, 2019 and December 31, 2021. Any potential bias that might be introduced by discarding the increasing number of S2 cells in order to protect privacy due to the decreasing number of travellers under travel restrictions was accounted for. For comparability of changes in mobility across districts, aggregated flows were further standardised using pre-pandemic mean baseline levels of mobility for the first eight weeks of 2020 (December 29, 2019 – February 22, 2020) (**SI Figure S2 – S4 & Figure S29**). This dataset was analysed by researchers at the University of Southampton, UK as per the terms of

149 the data sharing agreement. Production of this anonymized and aggregated dataset has been
150 detailed in previous studies.[3,37-39]

151 ***Stringency of COVID-19 Intervention***

152 Stringency Index of COVID-19 intervention policy in India data were obtained from the
153 Oxford COVID-19 Government Response Tracker (OxCGRT) project at state level and daily
154 temporal resolution (**SI Figure S5 & S30**). The Stringency Index is a composite index of
155 government responses to the COVID-19 pandemic compiled by OxCGRT based on data
156 collected from publicly available sources such as news articles, and government press releases
157 and briefings from 1 January 2020.[40,41] The project tracks national government policies and
158 interventions across a standardized series of indicators and creates a suite of composite indices
159 to measure the extent of these responses to understand how government responses evolved over
160 the course of the pandemic.[41] The Stringency Index was calculated as a composite score of 18
161 indicators of closure and containment, health, and economic policy.[24,40] Scores were created
162 using an additive unweighted approach, taking the ordinal value and adding a weighted
163 constant if the policy was general rather than targeted. The maximum values were rescaled to
164 create a score ranging from 0 to 100, with higher scores indicating stricter measures.[40]
165 Stringency index data for India were obtained from 27th April 2020 to 25th July 2021.

166 ***Climate data***

167 Three-dimensional Network Common Data Form (NetCDF) climate data were obtained from
168 the Copernicus Climate Data online repository (Copernicus Climate Change Service, Climate
169 Data Store, (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus
170 Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47
171 (Accessed on 19-05-2023). Data were ERA5 daily reanalysis global climate data obtained for
172 January 2019 to March 2021, gridded to 0.25 degrees of latitude and longitude. Variables
173 obtained were mean temperature of air (°C at 2m above the surface of land, sea or inland

174 waters), accumulated precipitation (metres), relative humidity (%) and downward ultraviolet
175 (UV, $KJ/m^2\ per\ hour$) radiation at the Earth's surface (**SI Figures S6-S9 & S31 – S34**).
176 ERA5 data is the fifth generation of European Centre for Medium-Range Weather
177 Forecasts (ECMWF) reanalysis for the global climate and weather for the past 4 to 7 decades.
178 Reanalysis is a method of combining model data with global observations for producing
179 complete and consistent datasets for a large number of atmospheric, ocean-wave and land-
180 surface quantities. Reanalysis works in the same way as the principle of data assimilation which
181 combines previous forecasts with newly available observations on a 12-hour basis to produce
182 new best estimates of atmospheric measures.[42] Climate data were extracted from NetCDF files
183 using the ncdf4[40,43] and RNetCDF[44] packages in R statistical software version 4.1.0 and
184 aggregated to district level using Quantum Geographic Information Systems (QGIS)
185 software.[45]
186 ***Urban and rural classification***
187 Data on the degree of urban, rural and suburban spatial area within each district (administrative
188 level II) were derived from the Global Human Settlement Layer (GHSL)[46] using the Degree of
189 Urbanisation – Territorial units classifier (GHS-DU-TUC) tool. The GHS-DU-TUC tool
190 classifies local units from a settlement classification grid according to the Degree of
191 Urbanisation (DEGURBA). It operationalises the method recommended by the 51st Session of
192 the United Nations Statistical Commission to delineate cities, urban and rural areas (stage
193 2, units classification) as defined by the Degree of Urbanisation Level 1 and 2. Categorised
194 variables for each degree of urbanisation (DEGURBA_L1_1 to DEGURBA_L1_3) were
195 generated for degree of urban vs. rural spatial area in each district area in accordance with
196 methods for implementation of INLA models outlined in Lezama-Ochoa *et al.* 2020.[47]
197 Degree of urbanisation was categorised as follows: (1) Rural (mostly thinly populated
198 areas), (2) Suburban (mostly intermediate density areas), and (3) Urban (mostly densely

populated areas). Population data for 2020 were obtained at 100m spatial resolution from the WorldPop online repository (https://www.worldpop.org/) and aggregated to calculate population density per km2 for each district. Data on public holidays time periods were obtained from the National Portal of India online repository (https://www.india.gov.in/). Public holidays, which included the date of public holiday and one day before and after, were assigned a value of 1. All other days were given a value 0.

Data analysis

Exploring changes in mobility in India during the pandemic

To gain a better understanding of travel networks and connectivity across India, we explored the overall patterns in domestic travel by rural, semi-urban and urban delineated areas in India, using weekly Google mobility data from November 10, 2019, to December 31, 2021. The relative levels of mobility across regions (regions are defined as six zones comprising different states in India defined under the States Reorganisation Act 1956[48]) and weeks were further calculated for each type of flow, relative to the mean level of pre-pandemic baseline in each region from December 29, 2019, to February 22, 2020. We also defined mobility reductions and communities of population movements between administrative level II units, i.e. districts, across the country for five periods (**SI Figure S2 - S3**): 1)

Pre-pandemic period (15 weeks) from November 10, 2019 to February 22, 2020; 2) First lockdown (6 weeks), from March 22 to May 2, 2020, that included strict travel restrictions, stay-at home orders and closure of many businesses; 3) Pre-second lockdown period (8 weeks) from January 31 to March 27, 2021; 4) Second lockdown (6 weeks) for the Delta wave, from April 18 to May 29, 2021; 5) post-second lockdown period (8 weeks), from November 7 to December 31, 2021, after travel restrictions for COVID-19 had been lifted in India. In the context of travel networks, a community refers to a group of areas that are more closely connected internally than with other areas in the network.[49,50] Community structures were

224 detected using the Louvain algorithm, a method of extracting communities from large
225 networks.[49] We mapped the communities identified to highlight distinct geographic groupings
226 of districts in terms of movements across periods.

227 ***Reproduction number***

228 To account for variations in the transmissibility of COVID-19, we estimated the instantaneous
229 reproduction number (R_t) for each district of the country with available case data (**SI Figure**
230 **S10 & S35**). First, the number of daily new COVID-19 cases at district level were smoothed
231 using a Gaussian smoothing approach over a 7-day rolling window.[51] Second, the mean
232 incidence of cases at day t was assumed following the Poisson distribution that is defined as:

233 $$E(I_t) = R_t \sum_{k=1}^{t} I_{t-k} w_k$$

234

235 where I_{t-k} is the incidence at time $t-k$, w_k is the infectivity profile which depends on the
236 serial interval of COVID-19 (5.2, 95%CI: 4.9–5.5).[52] The serial interval represents the time
237 between onset of the primary case to onset of the secondary case. Last, we estimated the daily
238 R_t for each district with a 7-day sliding window, using the EpiEstim package[53] in R statistical
239 software version 4.1.0.[54]

240 In order to account for changing transmissibility of COVID-19 caused by different
241 variants in the modelling, we also estimated the instantaneous basic reproduction number (R_0)
242 over time to capture the intrinsic transmission capability of the virus without interventions. We
243 first assembled data of the biweekly proportion of sequences of six main SARS-CoV-2
244 variants, including lineages B.1.1.7 (VOC Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2
245 (Delta), B.1.525 (Eta), and B.1.617.1 (Kappa), based on SARS-CoV-2 sequence data in the
246 Global Initiative on Sharing All Influenza Data (GISAID),[55] as of 25 October 2021. Using an
247 approach described by Ge *et al*.,[56] we then calculated a weighted average of basic reproduction

numbers of the six variants mentioned above and the SARS-CoV-2 strain in circulation before VOCs became predominant (seven coronavirus variants in total).

Models for examining lag-response associations between COVID-19 transmission and different factors

We built spatiotemporal Bayesian hierarchical models which consisted of weekly changes in the R_t of COVID-19 transmission for 665 India districts where data were available during 17 weeks from March 7 to July 3, 2021 (Delta wave) and during the 19 weeks between July 19th 2020 and November 29th 2020 (wave 1). We assumed that R_t adjusted by R_0, denoted as $\Delta R_t = R_t/R_0$, conformed to the Gamma distribution, $\Delta R_t | \mu\_t \sim Gamma(\frac{\mu_t}{0.5}, 0.5)$, where μ_t was the corresponding distribution expectation (or mean), reflecting the shape-rate parameterisation of the Gamma distribution used by the INLA package. Spatial and temporal fixed effects were accounted for in the model by incorporating terms for district and week, representing the spatial resolution of the data, and time scale during which data was collected. Spatiotemporal random effects were included to account for unobserved and unmeasured sources of variation in transmission and spatial and temporal dependency structures. First, for the expectation of ΔR_t within each city i, we constructed a base model below with two spatiotemporal random effects, r_t and b_i, and a fixed effect $v_{i,t}$

$$\mu_{i,t} = 1 + r_t + b_i + v_{i,t}$$

where $\Delta r_t = r_t - r_{t-1} \sim N(0, \tau^{-1})$ is a random walk model of order 1 (rw1), used to account for data temporality; b_i is a modified Besag-York-Mollie (BYM2) model for space, used to account for spatial variation across districts in the data; and the fixed effect $v_{i,t}$ of the cumulative infection rate among population was also included in the base model, as it might be related to herd immunity acquired by natural infection in previous waves before mass vaccination.

Second, as the evolution of COVID-19 is a complex process, and factors mentioned above might not be the only explanatory variables for the observed changes in transmission, we further examined the duration of public holidays as a fixed effect in models. All covariates obtained at daily temporal resolution were averaged by week. To account for multicollinearity of factors, we calculated pair-wise Pearson correlations for these variables and the variance inflation factor (VIF) for candidate variables in linear regressions for the whole country (**SI Figure S11 & S36)**. In order to account for any non-normally distributed data, we also calculated Kendall rank coefficients between explanatory variables in our model as a non-parametric exploration of multicollinearity (**SI Figure S12 & S37**). Estimations of multicollinearity were broadly similar using Pearson and Kendall rank correlation coefficients, with weaker associations found using Kendall rank coefficients. Collinear variables were therefore excluded based on the more conservative Pearson correlation coefficients. Variables with the highest VIF score and Pearson correlation coefficients of 0.5 which were excluded included relative humidity and UV, and only variables with a VIF score of less than 2.5 were retained. The relative impacts of remained factors, thus, was defined as the contributed percentage change in ΔR_t .

We built models of increasing complexity by systematically incorporating combinations of mobility, temperature, precipitation, stringency of intervention policy and public holidays covariates into our base model. Model goodness of fit was assessed using the deviance information criterion (DIC) and logarithmic score (logscore), consistent with previous studies[57], and final models for each pandemic wave selected. DIC balances model accuracy against complexity by estimating the number of effective parameters, while the logarithmic scores measure the predictive power of the model when excluding one data point at a time, with smaller values for each denoting better fitting models.

Third, we used the distributed lag non-linear models (DLNMs) formulation by defining lagged model covariates and a cross-basis matrix and incorporating the resulting cross-basis functions into our Bayesian spatiotemporal modelling framework. Using this approach we explored exposure-lag response associations between the relative risk (RR) of increase R_t in COVID-19 transmission, and changes in mobility, meteorological variations, and Stringency Index of intervention policy. DLNMs are a family of models that describe the lagged relationship between exposure and response variables in a model across both spatial and temporal dimensions.[58] DLNM models incorporate cross-basis functions that combine a lag-response function of variables at the temporal dimension and an exposure-response function to present the potential non-linear relationship along with the change of one factor. The resulting bi-dimensional exposure-lag-response function flexibly estimates the intensity of factors at varying time-lags after exposure.[58]

Given the common delays from infection to diagnosis and reporting, the lag-response impact of different factors on COVID-19 transmission were assessed by 0-3 weeks, with natural cubic splines selected for both the exposure and the lag dimensions. Last, we tested 18 candidate models of increasing complexity (with regard to input variables and model structure) with DLNMs for the whole country, and rural, suburban, and urban areas, respectively (**SI Table S2)**. DLNM cross-basis functions were built using R packages 'dlnm' and 'splines' and model parameters were estimated using the Integrated Nested Laplace Approximation (INLA) approach in R version 4.1.0. [54,60] INLA approaches include a wide and flexible class of models ranging from generalized linear mixed models to spatial and spatiotemporal models that are less computationally intensive therefore avoiding problems with model convergence.[59-61]

Finally, as no informed prior distribution estimates were available at the time of analyses, we explored the sensitivity of the best fit model to a range of uninformative priors. We specified a range of priors around the hyperparameters, i.e., τ, θ_1, and θ_2, in our base

model. Prior distributions were investigated for the best fit model using data for the Delta wave time period (**SI Table S4)** and for the wave 1 time period (**Tables S7**) using the deviance information criterion (DIC). The choice of prior distributions applied to best fit models using data from both waves was found to elicit only negligible measurable differences in model hyperparameters and DIC. Therefore, the prior used in this study was a penalized complexity prior with the precision $t = 1 / \sigma^2$, so that $\Pr(1/\sqrt{t} > 0.5) = 0.01$.

Model performance and validation

Model goodness-of-fit was assessed using DIC scores to compare model performances and identify the best-fitting model for the whole country, and rural, suburban, and urban areas, respectively. We also calculated the difference in mean absolute error (MAE) between the baseline model and the final selected model for each pandemic wave in order to identify the proportion of districts in different regions of India for which a more complex data-driven model improved model fit. Cross-validations using a leave-one-week-out and leave-one-state out approach were conducted to refit the selected model. This approach excluded one week or one state, respectively, from the fitting process during each cross-validation model iteration. Comparisons were made between observations and out-of-sample posterior predictive R_t for state and each week of both pandemic waves investigated. In order to validate DLNM model results, based on the findings of lag-response associations from analyses above, we incorporated lag-adjusted covariates into our spatiotemporal Bayesian hierarchical modelling and compared results with observations obtained from Bayesian spatiotemporal models incorporating DLNM models built using cross-basis functions.

Results

Spatiotemporal heterogeneity of mobility changes in India during the pandemic

Compared with baseline mean mobility patterns during the first 8 weeks of 2020, domestic travel within India dropped dramatically after the COVID-19 pandemic was declared by the

346 WHO and the country implemented its first lockdown for transmission containment (**Figure**
347 **1**). The lowest mobility level for domestic travel (26.9% of the pre-pandemic mean level) was
348 observed at week 15 of 2020 (April 5 – 11, 2020). In June 2020, restrictions on opening
349 shopping centres, religious places, hotels, and restaurants were lifted [32], coinciding with
350 increased population flows and an increase in infection cases. Overall, mobility gradually
351 recovered from mid-May 2020 to early March 2021, even during the first wave of COVID-19
352 in the second half of 2020.

353 Following a surge in transmission in March 2021, and concern about increased
354 infections and deaths caused by the Delta variant, another lockdown was implemented across
355 the country from mid-April to early June 2021. Domestic mobility during the second lockdown
356 reduced significantly from an average level of 90.5% in the 8 weeks between January 31 –
357 March 27, 2021, reaching its lowest level (54.6%) at week 20 of 2021 (May 16 – 22). However,
358 the stringency, compliance and duration of mobility reductions were less strict and shorter than
359 those of the first wave. Additionally, changes in mobility between rural, suburban and urban
360 districts of India displayed similar temporal patterns (**Figures 1C**), but travel in urban areas
361 (73.9%) was more affected by the pandemic compared with mobility in semi-urban (91.7%)
362 and rural (94.4%) areas in 2020 – 2021. In terms of geographic groupings of districts in
363 connected travel networks, there was also apparent spatiotemporal heterogeneity in response
364 to efforts to mitigate the varying scale of transmission across districts at different time periods
365 (**Figure 2**). During the pre-pandemic period from November 10, 2019, to February 22, 2020,
366 districts highly connected with each other formed 23 communities, with the 13 largest
367 communities containing 94.4% of districts in the country, ranging in size from 23 to 97 districts
368 per community.

369 Connections between districts in terms of travel were disrupted to mitigate transmission
370 during the first lockdown in 2020 (**SI Figure S3A**), and more isolated communities (n=79)

were formed with 54.4% (43) of them containing only one district. However, due to the fewer reductions in mobility (**SI Figure S3C**), connections between districts (31 communities) during the second lockdown were not as sparse as during the first lockdown. A total of 13 major communities contained 93.8% of districts, with a similar geographical range during the pre-pandemic period, except for some remote rural and semi-urban areas. Following the recovery of travel in November – December 2021 across the country (**SI Figure S3D**), connections between districts formed 22 communities, a pattern closely resembling that of pre-pandemic connected district communities.

Fig 1. COVID-19 cases, reproduction numbers and mobility changes in India during the pandemic. (**A**) Number of daily new confirmed COVID-19 cases reported in India from March 15, 2020, to December 25, 2021. (**B**) Estimated mean and 95% confidence interval (CI) of the basic reproduction number (R_0) and instantaneous reproduction rate (R_t). (**C**) Relative weekly mobility of domestic travel by rural, suburban and urban areas in India as measured by the aggregated Google COVID-19 mobility research dataset. Relative mobility levels were standardized by the overall mean level of each type of flow in each region during the first 8 weeks of 2020. The red and grey vertical dashed lines indicate the date of the COVID-19 pandemic being declared by the WHO and the first date of each year, respectively.

Fig 2. Changes in community domestic travel networks of Indian districts across four time periods in 2019-2021). (**A**) Communities (n=23) of domestic travel at district level during the pre-pandemic period from November 10, 2019, to February 22, 2020*. (**B**) Communities (n=79) of domestic travel during the first lockdown on March 22 - May 2, 2020*. (**C**) Communities (n=31) of domestic travel during the second lockdown on April 18 - May 29, 2021*. (**D**) Communities (n=22) of domestic travel post-second lockdown period (8 weeks), from November 7 to December 31, 2021, after travel restrictions for COVID-19 had been lifted in India. In each panel, geographically adjacent areas of the same colour represent an internally and closely connected community in terms of human movement in India. The community structure was detected using the Louvain algorithm. Circle size represents the relative volume of outbound travellers. The bigger the circle, the higher the level of outflow.* Based on aggregated Google COVID-19 mobility research dataset.

Nonlinear and lag-response impacts of mobility and other factors on the Delta wave

Upon initial exploratory analysis using baseline Bayesian spatiotemporal models for India, we found that population density as a fixed effect was not statistically associated with ΔR_t for COVID-19 transmission (DIC = 4756; **SI Table S2**). Humidity and UV were removed due to

a higher VIF and potential multicollinearity. We then included DLNMs for mobility, temperature, precipitation and Stringency Index, lagged between 0 and 3 weeks, including the holiday variable as a fixed effect in different candidate models (**SI Table S2**). The inclusion of mobility, temperature, precipitation and Stringency Index as DLNMs (Model 4.1) for the whole country and urban areas resulted in a greater reduction in the DIC and mean logarithmic score compared with the baseline model (**SI Figure S18**).

However, models which included DLNMs for mobility, temperature, and Stringency Index (Model 3.1) had the smallest DIC and logarithmic score in semi-urban areas. The best fitting model for rural areas only contained Stringency Index as DLNMs, which might be due to the smaller reductions in mobility in rural districts. The posterior predictive results from the best fitting model by cross validation showed that the model had a robust performance compared to observed data (**SI Figure S13 - S16**). The spatial random effects and the fitted Rt for the whole country were also presented in Supplementary (**SI Figures S17 - S20**). Based on results of the best fitting models for the whole country, overall, the recovery of mobility to 99% of the pre-pandemic level and the decrease of intervention stringency below 68 significantly increased the RR (>1) of COVID-19 transmission in India during the study period (**Figures 3A and 3J**). The increase of weekly precipitation (>0.15m) and cold weather (<27.2°C) were also associated with a higher risk of transmission (RR>1), but the variation in RRs in response to precipitation was small and the impact of temperature below 0°C was not significant (**Figures 3D and 3G**).

Fig 3. The lagged impact of different factors and scenarios on COVID-19 transmission during the Delta wave in 2021. (**A**) The overall association between mobility changes and COVID-19 transmission dynamics under 0- to 3-week lags. The red/blue lines show RR under the scenario of mobility below/above the overall mean level (0.99). The histogram with the secondary y-axis shows the frequency of data under different levels. (**B**) Contour plot of the association between mobility and relative risk (RR) of COVID-19 transmission. The deeper the shade of purple, the greater the increase in RR of transmission, while the deeper the shade of green, the greater the decrease in RR. (**C**) COVID-19 lag–response association for mobility level at 0.6, 0.8, 1.25, relative to the overall pre-pandemic mean level (1). The mean and 95% CI were presented. (**D**) – (**F**) Lag-response association between COVID-19 transmission and temperature (Temp) for cool (10°C), warm (20°C), and hot (30°C)

weather, relative to the overall mean of 27.2°C. (**G**) – (**I**) COVID-19 lag-response association for precipitation (Prec) at 0.05, 0.5, and 1m, relative to the overall mean of 0.15m. (**J**) – (**L**) Lag-response association between COVID-19 transmission and the stringency of intervention policy at low (40), medium (60), and high (80), relative to the overall mean Stringency Index (68.2). Results are for the best fitting model with DLNMs (base model + mobility + temperature + precipitation + intervention policy; see SI Table S2) across the whole country.

Fig 4. The lag-response association between COVID-19 transmission and different factors in urban, suburban, and rural districts. (**A**) – (**D**) COVID-19 lag–response association for different levels of mobility, temperature (Temp), precipitation (Prec) and the stringency of intervention policy in urban areas, relative to the overall mean level. Results are for the best fitting model with DLNMs (base model + mobility + temperature + precipitation + intervention policy) in urban districts. (**E**) – (**G**) Lag–response association between the risk of COVID-19 transmission and different levels of mobility, temperature (Temp), and the stringency of intervention policy in semi-urban areas, based on the best fitting model with DLNMs (base model + mobility + temperature + intervention policy; see SI Table S2) in suburban districts. (**H**) COVID-19 lag–response association for the Stringency Index of intervention policy at low (40), medium (60), and high (80), based on the best fitting model with DLNMs (base model + intervention policy) in rural districts. The mean and 95% CI of RR for each level were presented.

In addition, given the reporting delays of cases after exposure (i.e. incubation period plus the lags from illness onset, diagnosis to reporting, normally 10 days with an interquartile range of 8 – 11 days,[62] we found that the introduction of DLNMs improved model adequacy statistics compared with the inclusion of factors with no lags, which proved the rationality and necessity of considering the lag-response effects in the modelling. The maximum associations of mobility reductions (**Figure 3B**; relative mobility >0.5 times baseline mobility associated with RR of <0.8)) and intervention policy (**Figure 3K**; Stringency Index <30 associated with RR <0.95) with changes in R_t of COVID-19 transmission were found at a lag of 2 weeks with precipitation having an apparent maximum impact at a 1 to 2-week lag. However, we also found an increasing/decreasing risk of transmission under cool/hot weather at one 1-week lag (**Figure 3F**; temperature of <20°C associated with RR >1).

Similar lag-response patterns between COVID-19 transmission and covariates at different levels were also found in urban, suburban, and rural districts (**Figure 4**). Mobility of <0.8 times that of baseline was associated with a decrease in RR <1 at a 0 and 1-week lag in

urban delineated districts (**Figure 4A**). Association of stringency of government interventions with COVID-19 RR exhibited some heterogeneity between urban, suburban and rural areas however, with a Stringency Index of <40 associated with a RR>1 in urban areas at a 1-week lag (Figure 4D), but exhibiting a 2-week lagged response in suburban and rural areas (**Figures 4G & H**). Based on findings and methods described above, we re-tested models (without DLNMs) using all covariates with a 2-week lag, i.e. two weeks before cases reported and Rt observed (**SI Table S3**). We found that the best fitting model at country level was similar to the previous Model 4.1 with DLNMs, but precipitation was replaced by UV radiation due to potential multicollinearity. The results from leave-one-week-out cross-validation showed the best fitting 2-week lag-response model could further improve the prediction of dynamics in Rt of COVID-19 transmission across India (**SI Figures S23-S27**).

Comparing lag-response impacts of different factors between waves

We also ran Bayesian spatiotemporal models with DLNMs and data for the wave in 2020 (19th July to 29th November 2020) to compare drivers of transmission during both pandemic waves in 2020 (initial transmission wave) and 2021 (Delta wave). Results of DLNMs exploring drivers of COVID-19 transmission during the first wave were consistent with those exploring associations of COVID-19 transmission during the Delta wave in India. Humidity and UV were removed from the analyses due to multicollinearity ascertained by higher VIF statistics (>=2.5). Based on best fit model statistics (lowest DIC and mean logarithmic score compared to baseline model) DLNMs which best fit the data were models which included mobility, temperature, precipitation and Stringency Index (Model 4.1; **SI Table S3**). Consistent with model validation using data for the Delta wave, cross validation showed robust model results when posterior predictive results for the wave in 2020 were compared with observed data (**SI Figure S38 – S41**).

Fig 5. The lagged impact of different factors and scenarios on COVID-19 transmission during the initial wave of COVID-19 transmission in the second half of 2020. (**A**) The overall association between mobility changes and COVID-19 transmission dynamics under 0- to 3-week lags. The red/blue lines show RR under the scenario of mobility below/above the overall mean level (0.99). The histogram with the secondary y-axis shows the frequency of data under different levels. (**B**) Contour plot of the association between mobility and relative risk (RR) of COVID-19 transmission. The deeper the shade of purple, the greater the increase in RR of transmission, while the deeper the shade of green, the greater the decrease in RR. (**C**) COVID-19 lag–response association for mobility level at 0.8, 1.2, 1.4 relative to the overall pre-pandemic mean level (1). The mean and 95% CI were presented. (**D**) – (**F**) Lag-response association between COVID-19 transmission and temperature (Temp) for cool (10°C), warm (20°C), and hot (30°C) weather, relative to the overall mean of 25°C. (**G**) – (**I**) COVID-19 lag-response association for precipitation (Prec) at 0.5, 1.5, and 2.5m, relative to the overall mean of 0.24m. (**J**) – (**L**) Lag-response association between COVID-19 transmission and the stringency of intervention policy at three different measures of stringency: 70, 75, and 80, relative to the overall mean Stringency Index (76.5). Results are for the best fitting model with DLNMs (base model + mobility + temperature + precipitation + intervention policy; see SI Table S2) across the whole country.

Model results using data for the whole country found that a rebound in mobility to 1.2 and 1.4 times the mobility of pre-pandemic levels results in an increase in RR (>1) with a lag-time of between one and two weeks (**Figure 5C**). A high Stringency Index (80) was found to be associated with a lower RR with a two-and-a-half-week lag (**Figure 5L**) with a reduction in RR (<1) observed over a Stringency Index of 75 (**Figure 5J**). Cold weather (10°C & 20°C) was associated with a higher RR with a decrease in RR observed at higher temperatures (30°C) (**Figures 5D and 5F**). An increase in weekly precipitation (>0.2m) was also associated with an increase in transmission risk (**Figure 5G**) with a lag-time increase of between 1 and 2 weeks, although higher levels of weekly precipitation (>2.5m) were associated with a decrease again in RR (<1) at a 2 to 3-week lag.

Discussion

Using a de-identified and aggregated Google COVID-19 mobility research dataset, derived from time- and space-explicit mobile phone data, our study identified connected communities of travel networks, and quantified changes in population movements across rural and urban districts in India over the course of the pandemic. Our modelling results showed that mobility

533 changes, together stringency of government interventions and climate factors had lagged-
534 response impacts on the risk of COVID-19 transmission. The first nationwide lockdown
535 between March and June 2020, together with a reduction in population mobility, appear to have
536 been the main drivers for a relatively low transmission wave of COVID-19 in India during the
537 first half of 2020.[21,63] Although the announcement of the lockdown had initially resulted in an
538 increase in population mobility, with workers mostly representing informal sectors travelling
539 interstate to return home,[64] the majority of people travelling were not infected and this
540 population mobility therefore had little impact on transmission.[5]

541 In early 2021, NPI restriction measures such as social distancing and mask-wearing had
542 been gradually eased due to a sense of COVID-19 clinical infections being mild,[8,65] and inter-
543 state and rural to urban human mobility was seen to be increasing.[21,65] This included mass
544 attendance of political rallies and religious festivals, such as the Hindu festival Kumbh Mela
545 in India's most populous state of Uttar Pradesh where hundreds of thousands of people gathered
546 at the banks of the River Ganges.[21,23,65] The modelling results presented here indicate that this
547 recovery of mobility in early 2021 to 99% that of pre-pandemic levels, together with lower
548 stringency of government interventions and emergence of the more transmissible Delta variant,
549 contributed to higher transmission of COVID-19 infection during the Delta pandemic wave.
550 This is consistent with previously published research which attributed the surge of COVID-19
551 in April 2021 to the emergence of the more transmissible Delta variant (B.1.617 lineage) and
552 dominance as the main circulating strain, as well as relaxation of non-pharmaceutical
553 interventions.[13,21,66]

554 The second lockdown with reduced travel frequency and contact rates among
555 populations also played a significant role in mitigating COVID-19 spread across districts and
556 transmission in communities in the country. Mobility patterns were inversely associated with
557 the national Stringency Index, with a relative drop in mobility below 50% associated with a

Stringency Index of 80, consistent with previous research which found that community mobility, based on Google location data, drastically fell after the lockdown was instituted. However, the impacts of mobility changes were not fully synchronized between rural and urban areas, and the effects of travel restrictions and other interventions in slowing down COVID-19 transmission hinged on the intensity of these measures in reducing R_t of new variants with a higher transmissibility. Model results showed differences in lagged associations of COVID-19 RR with Stringency Index between rural, semi-urban and urban districts[24] and this was reflected in urban vs. rural transmission dynamics between both pandemic waves. During the first wave of COVID-19 in India, transmission was higher in urban rather than rural settings and cases were spatially clustered throughout metropolitan areas and peri-urban areas.[34,67,68] Conversely, during the Delta pandemic wave in India, cases were observed to be spreading more in rural areas where access to healthcare can be more limited than in urban areas.[18]

Climate covariates (temperature and precipitation) were also found to have lag-response associations with COVID-19 transmission, although these effects appear to be very limited in terms of relative risk. Modelling results for the Delta pandemic wave found a decrease in temperature (<20°C) was associated with an increased relative risk, consistent with previous modelling studies exploring climate impacts on COVID-19 transmission in India[33], and an increase in precipitation (>2.5m) associated with a decreased relative risk, with a 1 to 2-week lagged impact. This is consistent with wave 1 modelling results which found a 1 to 2-week lagged association between cold weather and precipitation on an increase in RR of COVID-19 transmission. Previous studies have also observed significant associations between COVID-19 transmission and temperature, dew point, humidity, and wind speed (Spearman's correlation)[69]; a 1 °C rise in mean temperature associated with increase in the daily number of COVID-19 confirmed cases when mean temperature was below 3 °C (GAM)[51]; a positive correlation between the number of infections with long-term climatic records of temperature,

583 wind speed, solar radiation[32]; average temperature correlated with COVID-19 based on
584 Spearman's correlation[40]; and minimum temperature and average temperature correlated with
585 the spread of COVID-19 in New York city (Spearman's correlation).[70] These previous
586 approaches most commonly used correlation analyses however, and spatiotemporal analyses
587 had not been used to examine such associations prior to the research we present here.

588 The work we have presented therefore builds upon previous research exploring the
589 driving factors that led to the surge in COVID-19 transmission during the Delta pandemic wave
590 in India [15,23,27], while also presenting a number of novel factors not previously presented in the
591 literature. Firstly, to our knowledge this is the first study to explore inter-district mobility
592 patterns in India during the initial and Delta waves of COVID-19 transmission, relative to pre-
593 pandemic levels, delineated by urban, suburban and rural location. By investigating these
594 changes in human mobility using fine spatial resolution Google COVID-19 Aggregated
595 Mobility Research data we have demonstrated that a surge in population movement, together
596 with an easing of NPIs were the main contributors to the surge in transmission during the Delta
597 pandemic wave. We have also explored the spatiotemporal heterogeneities in drivers of
598 transmission at district level accounting for urbanisation, building upon previous research
599 exploring the association between state level urbanisation and COVID-19 transmission.[29]

600 Additionally, the Bayesian hierarchical modelling approach we used provides a flexible
601 framework for quantifying heterogeneities in spatiotemporal drivers of transmission during
602 both pandemic waves while allowing complex and nonlinear relationships within the data to
603 be captured[60]. The ability for Bayesian models to incorporate spatial and temporal
604 dependencies in the models is particularly useful in regions such as India with substantial
605 divergence between urban and rural areas[71,72]. Integrating novel DLNM models into the
606 Bayesian framework allowed us to quantify lagged, nonlinear associations of drivers of

transmission with COVID-19 incidence, to account for the incubation period from infection to onset of clinical infection and delays in reporting of infection.

While our findings represent a comprehensive understanding of the drivers of transmission during the initial and Delta waves of COVID-19 transmission in India, these results should be interpreted in light of several important limitations. First, the Google mobility data is limited to smartphone users who have opted into Google's Location History feature, which is off by default. These data may not be representative of the population as whole, and furthermore their representativeness may vary by location. Importantly, these limited data are only viewed through the lens of differential privacy algorithms, specifically designed to protect user anonymity and obscure fine detail. However, comparisons between mobility datasets have shown good agreement with Google Location History data and other commonly used mobility data sources for capturing population-level mobility patterns [73]. Moreover, comparisons across rather than within locations are only descriptive since these regions can differ in substantial ways.

Second, the accuracy of our models relied on accurate estimates of R_t derived from reported case data, and R_0 estimates were proportional to the contact rate and might vary according to the local situation. The quality of reported data likely differed across districts due to varying case definitions, testing and surveillance capacity across the country, with various underreporting rate and reporting delays. Third, the Stringency Index data at state level used in spatiotemporal analyses for districts was formulated to assess lockdown strictness and measure the political commitment and strictness of governmental policies. These data did not measure the effectiveness of a country's response or provide information on how well policies were enforced. A higher value of Stringency Index did not necessarily mean that a country's response was better than that of those with lower values.[24,40] Fourth, many other factors (e.g.

vaccination and prior infections) might also contribute to COVID-19 transmission, but our models did not specify the contributions of these factors.

To our knowledge, this is the first study to combine human mobility data with Stringency Index and climate data within a Bayesian spatiotemporal framework to compare drivers of transmission by urban, suburban and rural district over the course of the pandemic in India, and quantify the lagged impact of these drivers on COVID-19 transmission risk. With the frequency of emerging infection outbreaks likely to increase in an increasingly urbanised global society, with more extreme weather events and pronounced changes in climate, the spatiotemporal modelling approach presented here provides a valuable framework for understanding drivers of infection transmission.[74,75] Based on our approach, examining how the lagged impact of human mobility, interventions and climate vary by urban and rural environment in their contribution to infection transmission can provide valuable insights into the intervention strategies in the future.

Acknowledgments: We thank the researchers and organisations who generated and publicly shared the mobility, epidemiological, intervention, sequencing data, and analysing code used in this research. We also thank Ms. Xilin Wu for collecting genomic data and sharing code for R0 calculation and Dr. Wenbin Zhang and Dr. Edson Utazi for commenting on the modelling framework. This study was supported by the National Institutes of Health (R01AI160780) and the Bill & Melinda Gates Foundation (INV-024911). The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. The views expressed in this article are those of the authors and do not represent any official policy.

Data Availability Statement: The code and data used for the analysis described in this study is available at the following GitHub repository: https://github.com/e3cleary/COVID-19-INDIA.git. The Google COVID-19 Aggregated Mobility Research Dataset used for this study is available with permission from Google LLC. Ethical clearance for collecting and using secondary data in this study was granted by the institutional review board of the University of Southampton (No. 61865). All data were supplied and analysed in an anonymous format, without access to personal identifying information.

References

1. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Advances in parasitology* 2006; **62**: 293-343.
2. Hâncean M-G, Slavinec M, Perc M. The impact of human mobility networks on the global spread of COVID-19. *Journal of Complex Networks* 2020; **8**(6): cnaa041.
3. Lai S, Ruktanonchai NW, Carioli A, et al. Assessing the effect of global travel and contact restrictions on mitigating the COVID-19 pandemic. *Engineering* 2021; **7**(7): 914-23.
4. Wu X, Tian H, Zhou S, Chen L, Xu B. Impact of global change on transmission of human infectious diseases. *Science China Earth Sciences* 2014; **57**: 189-203.
5. Gupta R, Pal SK, Pandey G. A comprehensive analysis of COVID-19 outbreak situation in India. *MedRxiv* 2020.
6. Basu D, Salvatore M, Ray D, et al. A comprehensive public health evaluation of lockdown as a non-pharmaceutical intervention on COVID-19 spread in India: National trends masking state level variations. *medRxiv* 2020.
7. Abraham P, Aggarwal N, Babu GR, et al. Laboratory surveillance for SARS-CoV-2 in India: Performance of testing & descriptive epidemiology of detected COVID-19, January 22-April 30, 2020. *The Indian journal of medical research* 2020; **151**(5): 424.
8. Jain P, Prakash O, Nyayanit DA, et al. Identification of SARS-CoV-2 clusters from symptomatic cases in India. *The Indian Journal of Medical Research* 2020; **152**(1-2): 111.
9. Kumar N, Hameed SKS, Babu GR, et al. Descriptive epidemiology of SARS-CoV-2 infection in Karnataka state, South India: Transmission dynamics of symptomatic vs. asymptomatic infections. *EClinicalMedicine* 2021; **32**: 100717.
10. Laxminarayan R, Wahl B, Dudala SR, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* 2020; **370**(6517): 691-7.
11. Kumar P, Chander B. COVID 19 mortality: Probable role of microbiome to explain disparity. *Medical hypotheses* 2020; **144**: 110209.
12. Chatterjee B, Karandikar RL, Mande SC. The mortality due to COVID-19 in different nations is associated with the demographic character of nations and the prevalence of autoimmunity. medRxiv 2020.07. 31.20165696. *BMJ Yale doi: https://doi org/101101/202007* 2020; **31**.
13. Dhar MS, Marwal R, Radhakrishnan V, et al. Genomic characterization and Epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *medRxiv* 2021.
14. Cherian S, Potdar V, Jadhav S, et al. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. *bioRxiv* 2021.
15. Mlcochova P, Kemp S, Dhar MS, et al. SARS-CoV-2 B. 1.617. 2 Delta variant replication and immune evasion. *Nature* 2021: 1-8.
16. Jain VK, Iyengar KP, Vaishya R. Differences between First wave and Second wave of COVID-19 in India. *Diabetes & metabolic syndrome* 2021; **15**(3): 1047.
17. Asrani P, Eapen MS, Hassan MI, Sohal SS. Implications of the second wave of COVID-19 in India. *The Lancet Respiratory Medicine* 2021; **9**(9): e93-e4.
18. Thiagarajan K. Why is India having a covid-19 surge? : British Medical Journal Publishing Group; 2021.
19. Laxminarayan R, Vinay T, Kumar KA, Wahl B, Lewnard JA. SARS-CoV-2 infection and mortality during the first epidemic wave in Madurai, south India: a prospective, active surveillance study. *The Lancet Infectious Diseases* 2021.
20. Jha P, Deshmukh Y, Tumbe C, et al. COVID mortality in India: National survey data and health facility deaths. *Science* 2022; **375**(6581): 667-71.

21. Ranjan R, Sharma A, Verma MK. Characterization of the Second Wave of COVID-19 in India. *medRxiv* 2021.
22. Jain VK, Iyengar KP, Vaishya R. Differences between First wave and Second wave of COVID-19 in India. *Diabetes & Metabolic Syndrome* 2021.
23. Choudhary OP, Singh I, Rodriguez-Morales AJ. Second wave of COVID-19 in India: dissection of the causes and lessons learnt. *Travel Medicine and Infectious Disease* 2021; **43**: 102126.
24. Periyasamy AG, Venkatesh U. Population mobility, lockdowns, and COVID-19 control: an analysis based on google location data and doubling time from India. *Healthcare Informatics Research* 2021; **27**(4): 325-34.
25. Woskie LR, Tsai TC, Wellenius GA, Jha A. Early Impact of India's Nationwide Lockdown on Aggregate Population Mobility and COVID-19 Cases. *Available at SSRN 3631258* 2020.
26. Praharaj S, King D, Pettit C, Wentz E. Using aggregated mobility data to measure the effect of COVID-19 policies on mobility changes in Sydney, London, Phoenix, and Pune. *Findings* 2020.
27. Praharaj S, Han H. Human mobility impacts on the surging incidence of COVID‐19 in India. *Geographical Research* 2022; **60**(1): 18-28.
28. Gupta A, Banerjee S, Das S. Significance of geographical factors to the COVID-19 outbreak in India. *Modeling earth systems and environment* 2020; **6**(4): 2645-53.
29. Praharaj S, Vaidya H. The urban dimension of COVID-19 in India: COVID Outbreak and Lessons for Future Cities. *New Delhi* 2020.
30. Bhadra A, Mukherjee A, Sarkar K. Impact of population density on Covid-19 infected and mortality rate in India. *Modeling Earth Systems and Environment* 2021; **7**(1): 623-9.
31. Praharaj S, Kaur H, Wentz E. The Spatial Association of Demographic and Population Health Characteristics with COVID‐19 Prevalence Across Districts in India. *Geographical Analysis* 2023; **55**(3): 427-49.
32. Gupta A, Banerjee S, Das S. Significance of geographical factors to the COVID-19 outbreak in India. *Modeling earth systems and environment* 2020; **6**: 2645-53.
33. Manik S, Mandal M, Pal S, Patra S, Acharya S. Impact of climate on COVID-19 transmission: A study over Indian states. *Environmental Research* 2022; **211**: 113110.
34. Bhunia GS, Roy S, Shit PK. Spatio-temporal analysis of COVID-19 in India–a geostatistical approach. *Spatial Information Research* 2021: 1-12.
35. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* 2020; **20**(5): 533-4.
36. COVID19 India. https://www.covid19india.org/ (accessed October 2021).
37. Bassolas A, Barbosa-Filho H, Dickinson B, et al. Hierarchical organization of urban mobility and its connection with city livability. *Nature communications* 2019; **10**(1): 4817.
38. Wilson RJ, Zhang CY, Lam W, Desfontaines D, Simmons-Marengo D, Gipson B. Differentially private SQL with bounded user contribution. *Proceedings on privacy enhancing technologies* 2020; **2020**(2): 230-50.
39. Ruktanonchai NW, Floyd J, Lai S, et al. Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* 2020; **369**(6510): 1465-70.
40. Hale T, Angrist N, Kira B, Petherick A, Phillips T, Webster S. Variation in government responses to COVID-19. 2020.
41. Hale T, Angrist N, Goldszmidt R, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature human behaviour* 2021; **5**(4): 529-38.

42. Copernicus Climate Data Store. ERA5 hourly data on single levels from 1959 to present. https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview2021).
43. Pierce D, Pierce MD. Package 'ncdf4'. *See https://www vps fmvz usp br/CRAN/web/packages/ncdf4/ncdf4 pdf* 2019.
44. Michna P, Woods M. RNetCDF–A package for reading and writing NetCDF datasets. *The R Journal* 2013; **5**(2): 29-36.
45. QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. 2021. http://qgis.osgeo.org.
46. European Commission and Statistical Office of the European Union. Global Human Settlement Layer. https://ghsl.jrc.ec.europa.eu/tools.php (accessed July 2021.
47. Lezama-Ochoa N, Pennino MG, Hall MA, Lopez J, Murua H. Using a Bayesian modelling approach (INLA-SPDE) to predict the occurrence of the Spinetail Devil Ray (Mobular mobular). *Scientific reports* 2020; **10**(1): 1-11.
48. THE STATES REORGANISATION ACT. 1956.
49. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008; **2008**(10): P10008.
50. Li Z, Huang X, Ye X, Li X. ODT flow explorer: Extract, query, and visualize human mobility. *arXiv preprint arXiv:201112958* 2020.
51. Xie J, Zhu Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment* 2020; **724**: 138201.
52. Alene M, Yismaw L, Assemie MA, Ketema DB, Gietaneh W, Birhan TY. Serial interval and incubation period of COVID-19: a systematic review and meta-analysis. *BMC Infectious Diseases* 2021; **21**: 1-9.
53. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* 2013; **178**(9): 1505-12.
54. R Core Team. R: A language and environment for statistical computing. R Foundation for
Statistical Computing. 2021. https://www.R-project.org/.
55. Elbe S, Buckland‐Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global challenges* 2017; **1**(1): 33-46.
56. Ge Y, Zhang W, Liu H, et al. Effects of worldwide interventions and vaccination on COVID-19 between waves and countries. 2021.
57. Lowe R, Lee SA, O'Reilly KM, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *The Lancet Planetary Health* 2021; **5**(4): e209-e19.
58. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non‐linear models. *Statistics in medicine* 2010; **29**(21): 2224-34.
59. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* 2009; **71**(2): 319-92.
60. Moraga P. Geospatial health data: Modeling and visualization with R-INLA and shiny: CRC Press; 2019.
61. Martino S, Rue H. Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the inla program. *Department of Mathematical Sciences, NTNU, Norway* 2009.
62. Xi W, Pei T, Liu Q, et al. Quantifying the time-lag effects of human mobility on the COVID-19 transmission: a multi-city study in China. *Ieee Access* 2020; **8**: 216752-61.

63. Mave V, Shaikh A, Monteiro JM, Bogam P, Pujari BS, Gupte N. Impact of National and Regional Lockdowns on Growth of COVID-19 Cases in COVID-Hotspot City of Pune in Western India: A Real-World Data Analysis. *medRxiv* 2021: 2021.05. 05.21254694.
64. World Health Organization. Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection, 26 May 2020: World Health Organization, 2020.
65. Lodha R, Kabra S. Second COVID-19 Surge: Challenges and Handling. Springer; 2021.
66. Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *The Lancet* 2020; **395**(10233): 1382-93.
67. Murhekar MV, Bhatnagar T, Selvaraju S, et al. SARS-CoV-2 antibody seroprevalence in India, August–September, 2020: findings from the second nationwide household serosurvey. *The Lancet Global Health* 2021; **9**(3): e257-e66.
68. Murhekar MV, Bhatnagar T, Selvaraju S, et al. Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020. *Indian Journal of Medical Research* 2020; **152**(1): 48.
69. Şahin M. Impact of weather on COVID-19 pandemic in Turkey. *Science of the Total Environment* 2020; **728**: 138810.
70. Bashir MF, Ma B, Komal B, Bashir MA, Tan D, Bashir M. Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of the Total Environment* 2020; **728**: 138835.
71. Ahluwalia IJ, Kanbur R, Mohanty PK. Urbanisation in India: Challenges, opportunities and the way forward. 2014.
72. Bhagat RB. Emerging pattern of urbanisation in India. *Economic and political weekly* 2011: 10-2.
73. Ruktanonchai NW, Ruktanonchai CW, Floyd JR, Tatem AJ. Using Google Location History data to quantify fine-scale human mobility. *International journal of health geographics* 2018; **17**: 1-13.
74. Karn M, Sharma M. Climate change, natural calamities and the triple burden of disease. *Nature Climate Change* 2021; **11**(10): 796-7.
75. Guest edited by Zhi-Jie Zhang L-QF, Michael P. Ward, Ousman Bajinka, Wei Hu, and Xiao-Nong Zhou, . Guest edited by Zhi-Jie Zhang, Li-Qun Fang, Michael P. Ward, Ousman Bajinka, Wei Hu, and Xiao-Nong Zhou. *A thematic series in Infectious Diseases of Poverty* 2023.

Supporting information

Figure S1. Regions in India investigated by this study and the number and density of population at district level (administrative level II) in 2020. Areas shaded in grey are areas for which no data is available.

Figure S2. Five periods for travel network modularity analysis (A): 1) Pre-pandemic period (15 weeks) from November 10, 2019 to February 22, 2020; 2) First lockdown (6 weeks), from March 22 to May 2, 2020, that included strict travel restrictions, stay-at home orders and closure of many businesses; 3) Pre-

second lockdown period (8 weeks) from January 31 to March 27, 2021; 4) Second lockdown (6 weeks) for the Delta wave, from April 18 to May 29, 2021; 5) post-second lockdown period (8 weeks), from November 7 to December 31, 2021, after travel restrictions for COVID-19 had been lifted in India.

Fig S3. Relative changes of outbound travel from districts across India during the pandemic compared with average pre-pandemic levels during the 12 weeks from November 10, 2019, to February 22, 2020. (**A**) Reductions of outbound flows under the first lockdown during the 6-week period from March 22 to May 2, 2020. (**B**) Changes in outflow during the 8-week period from January 31 to March 27, 2021, before the second lockdown. (**C**) Reductions of outflows during the 6-week second lockdown from April 18 to May 29, 2021. (**D**) Changes in outflow during the 8-week period from November 7 to December 31, 2021. Sub-division maps at administrative level I (state) and II (district) were obtained from the GADM version 3.6 (https://gadm.org/). Regions in which outflow data are not available are those represented in green. Areas shaded in grey are areas for which no data is available.

Table S1. Summary Statistics for data used for wave 1 and Delta wave spatiotemporal models

SI Delta Wave

Table S2. Wave 2: Adequacy results for models with DLNMs and increasing complexity.

Table S3. Wave 2: Adequacy results for models (without DLNMs) using 2-week lag covariates with increasing complexity.

Table S4. Model hyperparameters using a range of prior distributions in best fit model 4.1 for Delta Wave

Figure S4. Relative intra-district mobility during the Delta wave in India, standardised by pre-pandemic mean baseline levels of mobility for the first eight weeks of 2020 (December 29, 2019 – February 22, 2020) for each district. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S5. Stringency Index of COVID-19 intervention policy implemented during the Delta wave in India. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S6. Mean temperature at 2m above the surface during the Delta wave in India. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S7. Accumulated weekly precipitation (metres) during the Delta wave in India. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S8. Relative humidity during the Delta wave in India. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S9. Downward ultraviolet (UV) radiation (KJ/m2 per hour) during the Delta wave in India. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S10. Weekly Rt derived from COVID-19 cases reported during the Delta wave in India. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S11. Pairwise Pearson correlations between weekly means of variables at district level during the Delta wave in India, 2021. R0: basic reproduction number. Rt: instantaneous reproduction number. ln_R: log(Rt/R0). Cases_rate: new COVID-19 cases reported per 1000 people. Cases_accu_rate: cumulative cases per 1000 people reported since the first week of the wave. mean_intra: intra-district relative mobility. d2m: relative humidity. t2m: mean temperature of air (°C at 2m above the surface of land, sea or inland waters). tp: precipitation (metres). uv: downward ultraviolet radiation. Stringency: index of COVID-19 intervention stringency. Holiday: days of public holidays in a week. pop_sum: total population of each district. pop_density: population number per km2 of each district.

Figure S12. Kendall rank correlations between weekly means of variables at district level during the Delta wave in India, 2021. R0: basic reproduction number. Rt: instantaneous reproduction number. ln_R: log(Rt/R0). Cases_rate: new COVID-19 cases reported per 1000 people. Cases_accu_rate: cumulative cases per 1000 people reported since the first week of the wave. mean_intra: intra-district relative mobility. d2m: relative humidity. t2m: mean temperature of air (°C at 2m above the surface of land, sea or inland waters). tp: precipitation (metres). uv: downward ultraviolet radiation. Stringency: index of COVID-19 intervention stringency. Holiday: days of public holidays in a week. pop_sum: total population of each district. pop_density: population number per km2 of each district.

Figure S13. Posterior predictive mean Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1) at country level using leave-one-week-out cross-validation approach. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S14. Standard deviation (SD) of posterior predictive Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1 without DLNMs) at country level using a leave-one-week-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Figure S15. Posterior predictive mean Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1) at country level using leave-one-state-out cross-validation approach. The weeks in 2021 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S16. Standard deviation (SD) of posterior predictive Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1 without DLNMs) at country level using a leave-one-state-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Figure S17. Contribution of spatial random effects to estimates of Rt changes in the base model. Areas shaded in grey are areas for which no data is available.

Figure S18. Improvement by using the best fitting model across the country, compared to baseline model. Difference between mean absolute error (MAE) for the baseline model (weekly random effects, spatial random effects and population density) and MAE for the best fitting model (model 4.1 with DLNMs). Districts with positive values (pink) suggest that capturing the nonlinear and delayed impacts of mobility, climate information and intervention stringency, improves the model in these areas. Districts with negative values (blue) suggest that mobility, intervention and climate information did not improve the model fit and other unexplained factors might dominate space-time dynamics in these areas. The MAE of the selected model was smaller than the baseline model for 385 of the 665 (57.9%) districts in India, with the results of model performance provided by geo-political regions in the Table. Areas shaded in grey are areas for which no data is available.

Figure S19. Observed versus posterior fitted Rt in the capital district of each state using the best fitting model (model 4.1 with DLNMs) at country level. Graphs with a log scale at y-axis show the observed Rt derived from reported case data, and corresponding mean and 95% confidence interval (CI, shaded pink area) of fitted Rt, derived from the best fitting model (model 4.1 with DLNMs) at country level. States are ordered by their geographical location.

Figure S20. Observed versus posterior predictive Rt in the capital district of each state, using leave-one-week-out cross-validation approach**.** Graphs with a log scale at y-axis show the observed Rt derived from reported case data, and corresponding posterior predictive mean and 95% prediction interval (CI, shaded pink area), derived from the best fitting model (model 4.1 with DLNMs) at country level. States are ordered by their geographical location.

Figure S21. Contribution of spatial random effects to estimates of Rt changes in the base model. Areas shaded in grey are areas for which no data is available.

Figure S22. Improvement of using the best fitting model with 2-week lag covariates (no DLNMs), compared to baseline model with the same lag. Difference between mean absolute error (MAE) for the baseline model and MAE for the best fitting model (Model 4.1). Districts with positive values (pink) suggest that capturing the 2-week lag impacts of mobility, temperature, UV and intervention stringency, improves the model in these areas. Districts with negative values (blue) suggest that mobility, intervention and climate information did not improve the model fit and other unexplained factors might dominate space-time dynamics in these areas. The MAE of the selected model was smaller than the baseline model for 428 of the 665 (64.4%) districts in India, and further improved the best fitting model with DLNMs (Figure S12). Results of model performance are provided by geo-political regions in the Table. Areas shaded in grey are areas for which no data is available.

Figure S23. Posterior predictive mean Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag covariates and leave-one-week-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Figure S24. Standard deviation (SD) of posterior predictive Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag covariates and leave-one-week-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Fig S25. Observed versus posterior predictive Rt in the capital district of each state. Graphs with a log scale at y-axis show the observed Rt derived from reported case data, and corresponding posterior predictive mean and 95% prediction interval (CI, shaded pink area), derived from the best fitting model without DLNMs at country level (model 4.1: base model + mobility + temperature + UV + intervention policy; see SI Table S2), using 2-week lag covariates and leave-one-week-out cross-validation approach. States are ordered by their geographical location.

Figure S26. Posterior predictive mean Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag covariates and leave-one-state-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Figure S27. Standard deviation (SD) of posterior predictive Rt during the Delta wave in India, 2021, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag

covariates and leave-one-state-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

SI Wave 1

Table S5. Wave 1: Adequacy results for models with DLNMs and increasing complexity.

Table S6. Wave 1: Adequacy results for models (without DLNMs) using 2-week lag covariates with increasing complexity.

Table S7. Model hyperparameters using a range of prior distributions in best fit model 4.1 for Wave 1

Figure S28. COVID-19 cases reported by district each week during wave 1 in India. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S29. Relative intra-district mobility during wave 1 in India, standardised by pre-pandemic mean baseline levels of mobility for the first eight weeks of 2020 (December 29, 2019 – February 22, 2020) for each district. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S30. Stringency Index of COVID-19 intervention policy implemented during wave 1 in India. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S31. Mean temperature at 2m above the surface during wave 1 in India. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S32. Accumulated weekly precipitation (metres) during wave 1 in India. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S33. Relative humidity during wave 1 in India. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S34. Downward ultraviolet (UV) radiation (KJ/m2 per hour) during wave 1 in India. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S35. Weekly Rt derived from COVID-19 cases reported during the wave 1 in India. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Figure S36. Pairwise Pearson correlations between weekly means of variables at district level during the wave 1 in India, 2020. R0: basic reproduction number. Rt: instantaneous reproduction number. ln_R: log(Rt/R0). Cases_rate: new COVID-19 cases reported per 1000 people. Cases_accu_rate: cumulative cases per 1000 people reported since the first week of the wave. mean_intra: intra-district relative mobility. d2m: relative humidity. t2m: mean temperature of air (°C at 2m above the surface of land, sea or inland waters). tp: precipitation (metres). uv: downward ultraviolet radiation. Stringency: index of COVID-19 intervention stringency. Holiday: days of public holidays in a week. pop_sum: total population of each district. pop_density: population number per km2 of each district.

Figure S37. Kendall rank correlations between weekly means of variables at district level during the wave 1 in India, 2020. R0: basic reproduction number. Rt: instantaneous reproduction number. ln_R: log(Rt/R0). Cases_rate: new COVID-19 cases reported per 1000 people. Cases_accu_rate: cumulative cases per 1000 people reported since the first week of the wave. mean_intra: intra-district relative mobility. d2m: relative humidity. t2m: mean temperature of air (°C at 2m above the surface of land, sea or inland waters). tp: precipitation (metres). uv: downward ultraviolet radiation. Stringency: index of COVID-19 intervention stringency. Holiday: days of public holidays in a week. pop_sum: total population of each district. pop_density: population number per km2 of each district.

Figure S38. Posterior predictive mean Rt during wave 1 in India, 2020, derived from the best fitting model (model 4.1) at country level using leave-one-week-out cross-validation approach. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Fig S39. Standard deviation (SD) of posterior predictive Rt during wave 1 in India, 2020, derived from the best fitting model (model 4.1) at country level leave-one-week-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Figure S40. Posterior predictive mean Rt during wave 1 in India, 2020, derived from the best fitting model (model 4.1) at country level using leave-one-district-out cross-validation approach. The weeks in 2020 investigated are numbered in maps. Areas shaded in grey are areas for which no data is available.

Fig S41. Standard deviation (SD) of posterior predictive Rt during wave 1 in India, 2020, derived from the best fitting model (model 4.1) at country level leave-one-district-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Figure S42. Contribution of spatial random effects to estimates of Rt changes in the base model. Areas shaded in grey are areas for which no data is available.

Figure S43. Improvement by using the best fitting model across the country, compared to baseline model. Difference between mean absolute error (MAE) for the baseline model (weekly random effects, spatial random effects and population density) and MAE for the best fitting model (model 4.1 with DLNMs). Districts with positive values (pink) suggest that capturing the nonlinear and delayed impacts of mobility, climate information and intervention stringency, improves the model in these areas. Districts with negative values (blue) suggest that mobility, intervention and climate information did not improve the model fit and other unexplained factors might dominate space-time dynamics in these areas. The MAE of the selected model was smaller than the baseline model for 430 of the 661 (65.17%) districts in India, with the results of model performance provided by geo-political regions in the Table. Areas shaded in grey are areas for which no data is available.

Figure S44. Observed versus posterior fitted Rt in the capital district of each state using the best fitting model (model 4.1 with DLNMs) at country level. Graphs with a log scale at y-axis show the observed Rt derived from reported case data, and corresponding mean and 95% confidence interval (CI, shaded pink area) of fitted Rt, derived from the best fitting model (model 4.1 with DLNMs) at country level. States are ordered by their geographical location.

Figure S45. Observed versus posterior predictive Rt in the capital district of each state, using leave-one-week-out cross-validation approach**.** Graphs with a log scale at y-axis show the observed Rt derived from reported case data, and corresponding posterior predictive mean and 95% prediction interval (CI, shaded pink area), derived from the best fitting model (model 4.1 with DLNMs) at country level. States are ordered by their geographical location.

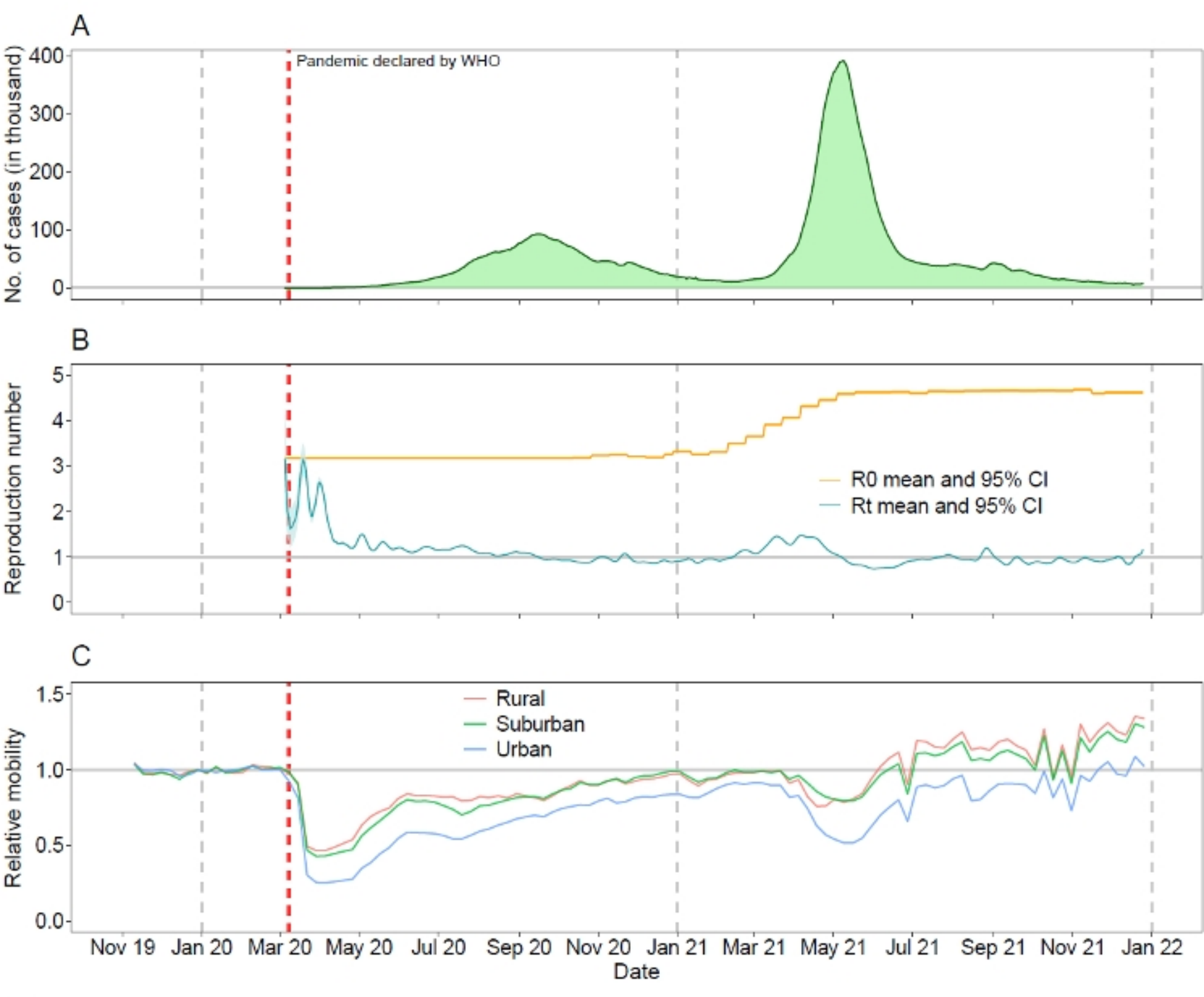
Figure S46. Posterior predictive mean Rt during the wave 1 in India, 2020, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag covariates and leave-one-week-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Fig S47. Standard deviation (SD) of posterior predictive Rt during wave 1 in India, 2020, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag covariates and

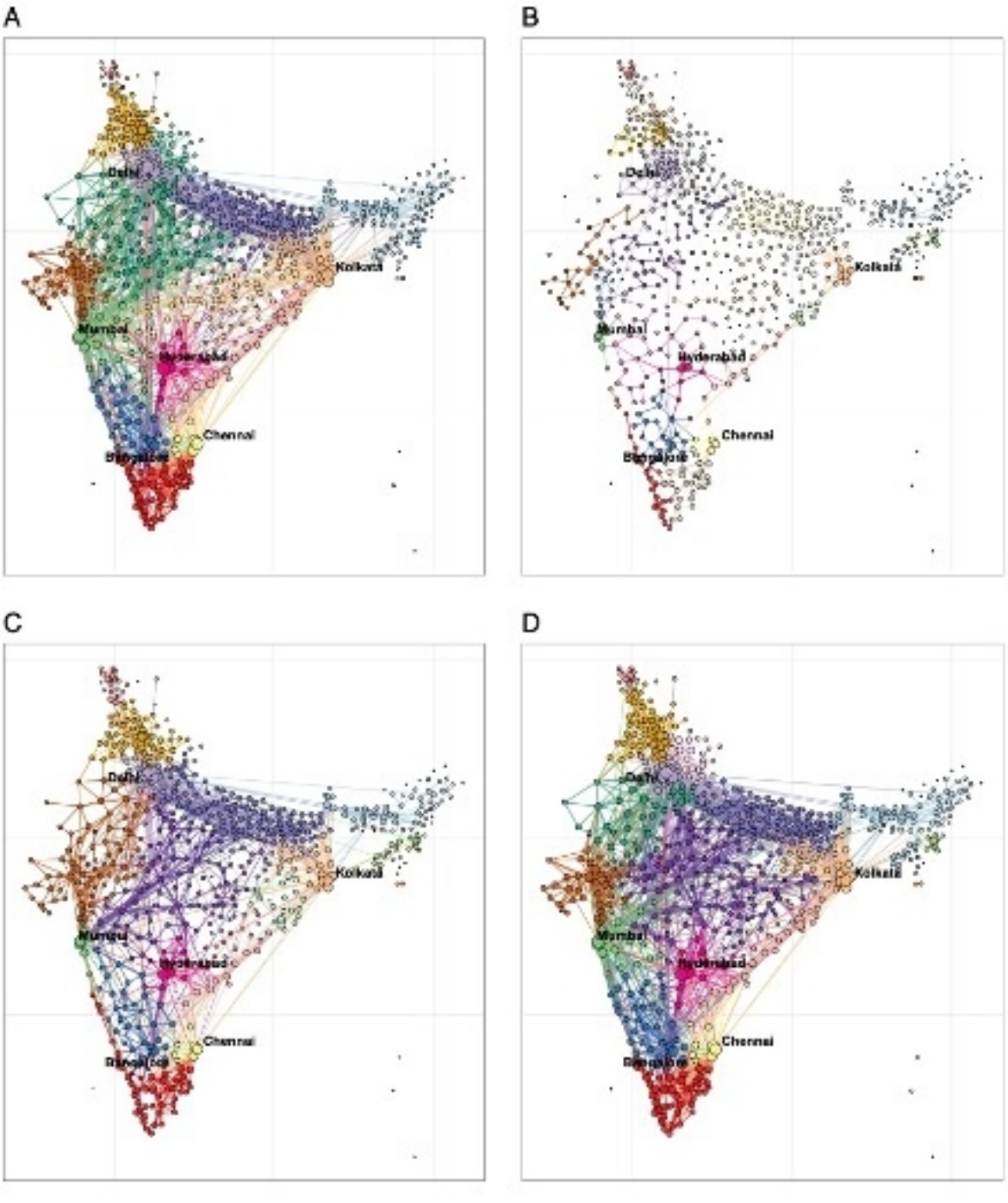
leave-one-week-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

Figure S48. Posterior predictive mean Rt during the wave 1 in India, 2020, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag covariates and leave-one-district-out cross-validation approach. Areas shaded in grey are areas for which no data is available.

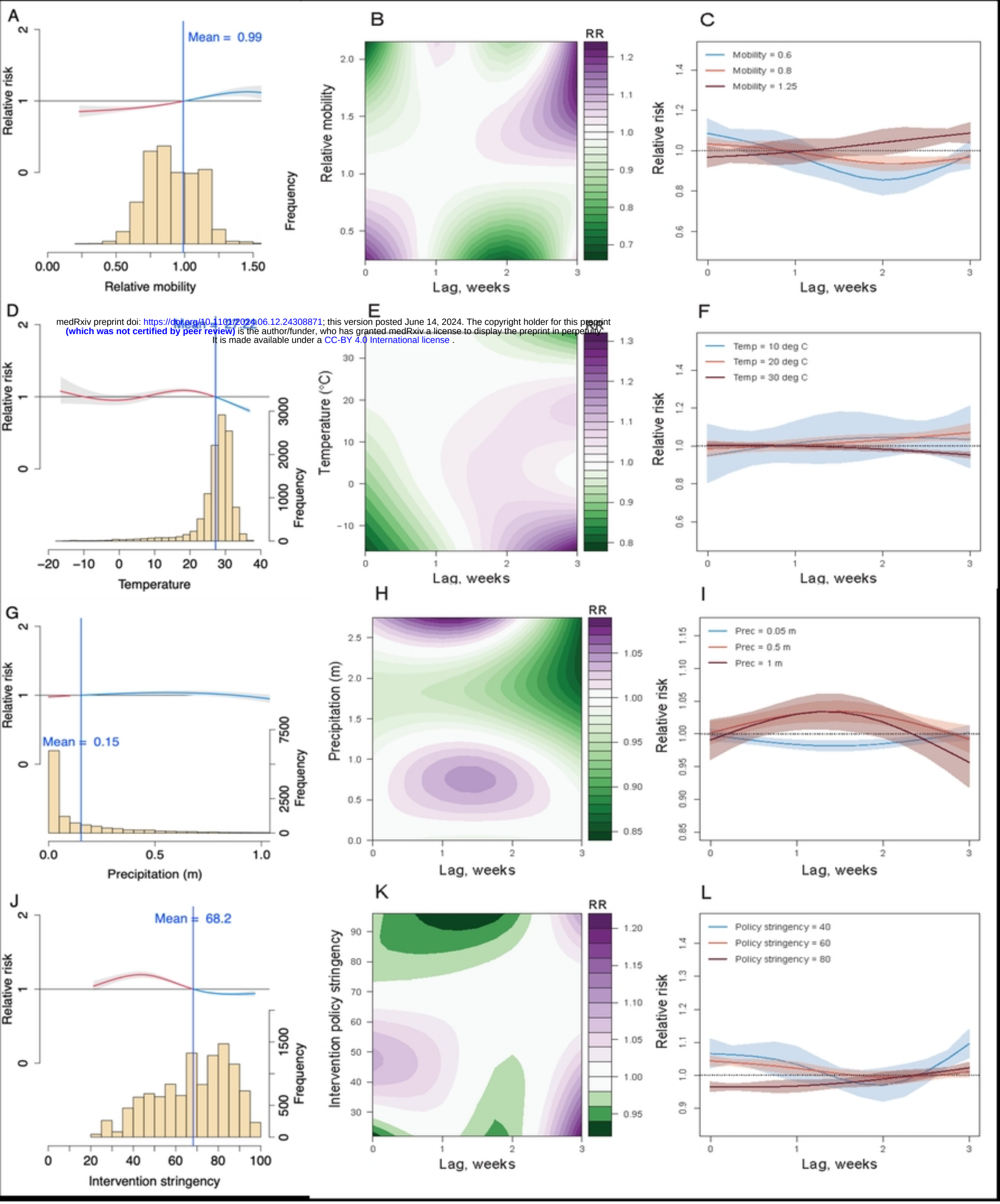
Fig S49. Standard deviation (SD) of posterior predictive Rt during wave 1 in India, 2020, derived from the best fitting model (model 4.1 without DLNMs) at country level using 2-week lag covariates and leave-one-district-out cross-validation approach. Areas shaded in grey are areas for which no data is available.


A
Pandemic declared by WHO
No. of cases (in thousand)
400
300
200
100
0
B
Reproduction number
5
4
3
2
1
0
R0 mean and 95% CI
Rt mean and 95% CI
C
Relative mobility
1.5
1.0
0.5
0.0
Rural
Suburban
Urban
Nov 19
Jan 20
Mar 20
May 20
Jul 20
Sep 20
Nov 20
Jan 21
Mar 21
May 21
Jul 21
Sep 21
Nov 21
Jan 22
Date

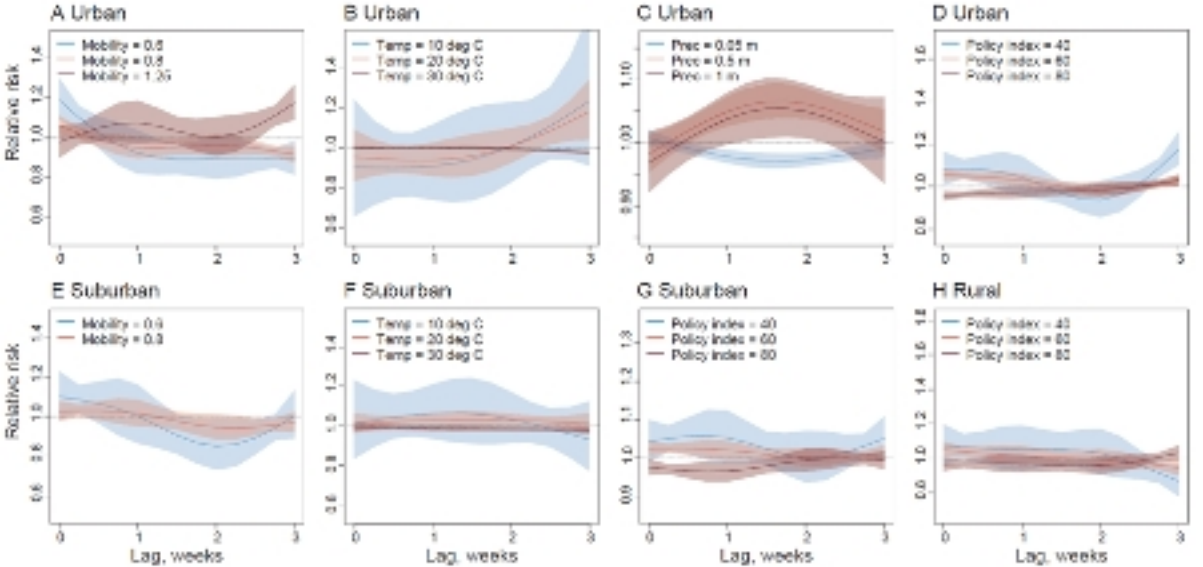
Figure


A
Delhi
Kolkata
Mumbai
Hyderabad
Chennai
Bangalore
B
Delhi
Kolkata
Mumbai
Hyderabad
Chennai
Bangalore
C
Delhi
Kolkata
Mumbai
Hyderabad
Chennai
Bangalore
D
Delhi
Kolkata
Mumbai
Hyderabad
Chennai
Bangalore

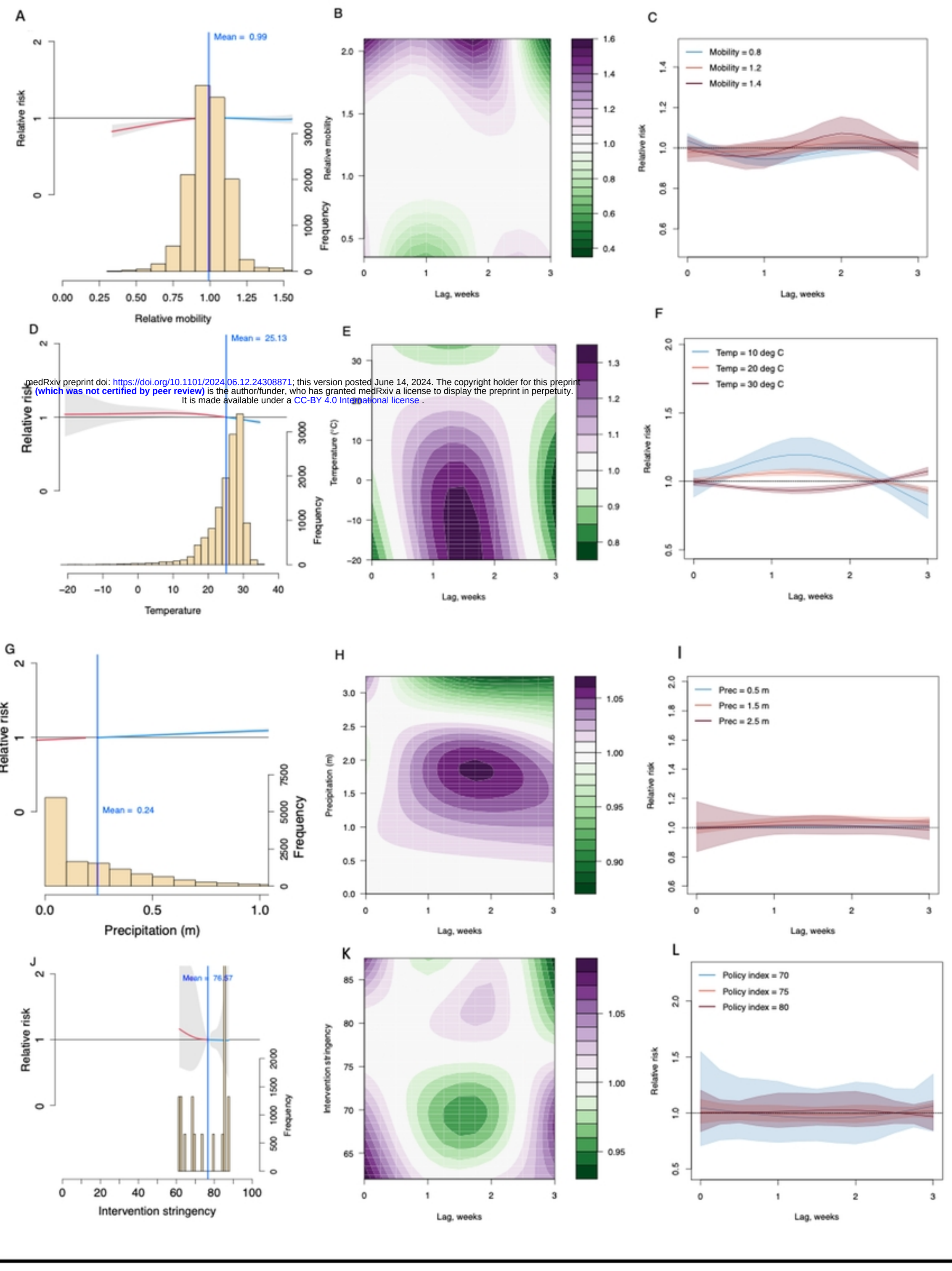
Figure



Figure


A Urban
B Urban
C Urban
D Urban
E Suburban
F Suburban
G Suburban
H Rural
Relative risk
Lag, weeks
Mobility = 0.6
Mobility = 0.8
Mobility = 1.25
Temp = 10 deg C
Temp = 20 deg C
Temp = 30 deg C
Prec = 0.05 m
Prec = 0.5 m
Prec = 1 m
Policy index = 40
Policy index = 60
Policy index = 80

Figure



Figure