Title: Impact of close interpersonal contact on COVID-19 incidence: evidence from one year of mobile device data

Short title: Close interpersonal contact predicts COVID-19 incidence

One sentence summary: Close interpersonal contact measured using mobile device location data explains dynamics of COVID-19 transmission in Connecticut during the first year of the pandemic.

Authors: Forrest W. Crawford[1,2,3,4], Sydney A. Jones[5,6], Matthew Cartter[6], Samantha G. Dean[1], Joshua L. Warren[1], Zehang Richard Li[7], Jacqueline Barbieri[8], Jared Campbell[8], Patrick Kenney[8], Thomas Valleau[8], Olga Morozova[9]

Affiliations:
1. Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA.
2. Department of Statistics & Data Science, Yale University, New Haven, CT, USA.
3. Department of Ecology & Evolutionary Biology, Yale University, New Haven, CT, USA.
4. Yale School of Management, New Haven, CT, USA.
5. Epidemic Intelligence Service, Centers for Disease Control & Prevention, Atlanta, GA, USA.
6. Infectious Diseases Section, Connecticut Department of Public Health, New Haven, CT, USA.
7. Department of Statistics, University of California, Santa Cruz, Santa Cruz, CA, USA.
8. Whitespace Solutions, Ltd, Alexandria, VA, USA.
9. Program in Public Health and Department of Family, Population and Preventive Medicine, Stony Brook University, NY, USA.

Abstract: Close contact between people is the primary route for transmission of SARS-CoV-2, the virus that causes coronavirus disease 2019 (COVID-19). We sought to quantify interpersonal contact at the population-level by using anonymized mobile device geolocation data. We computed the frequency of contact (within six feet) between people in Connecticut during February 2020 – January 2021. Then we aggregated counts of contact events by area of residence to obtain an estimate of the total intensity of interpersonal contact experienced by residents of each town for each day. When incorporated into a susceptible-exposed-infective-removed (SEIR) model of COVID-19 transmission, the contact rate accurately predicted COVID-19 cases in Connecticut towns during the timespan. The pattern of contact rate in Connecticut explains the large initial wave of infections during March–April, the subsequent drop in cases during June–August, local outbreaks during August–September, broad statewide resurgence during September–December, and decline in January 2021. Contact rate data can help guide public health messaging campaigns to encourage social distancing and in the allocation of testing resources to detect or prevent emerging local outbreaks more quickly than traditional case investigation.

Keywords: SARS-CoV-2, infectious disease, transmission, mobility

Introduction

Close contact between people is the primary route for transmission of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease (COVID-19) [1]. Social distancing guidelines published by the United States (U.S.) Centers for Disease Control and Prevention (CDC) recommend that people stay at least six feet away from others to avoid transmission via direct contact or exposure to respiratory droplets [2]. Throughout the world, non-pharmaceutical interventions, including social distancing guidelines and stay-at-home orders, have been employed to encourage the physical separation of people and reduce the risk of COVID-19 transmission via close contact [3–6]. U.S. states with the lowest levels of self-reported social distancing behavior are experiencing most severe COVID-19 outbreaks [7].

While individual-level compliance with social distancing guidelines can be difficult to measure, researchers have proposed population-level mobility metrics based on mobile device geolocation data as a proxy measure for physical distancing and movement patterns during the COVID-19 pandemic [8–12]. Investigators have characterized geographic and temporal changes in mobility metrics following non-pharmaceutical interventions like social distancing guidelines and stay-at-home mandates during the COVID-19 pandemic [11, 13–22]. Researchers have also studied the association between mobility metrics and COVID-19 cases or other proxy measures of transmission [11, 23–34]. Most mobility metrics measure aggregated movement patterns of individual mobile devices: time spent away from home, distance traveled, or density of devices appearing in an area during a given time interval. CDC reports mobility metrics from Google, Safegraph, and Cuebiq [35]. Some mobility metrics measure spatial relationships among individual devices. Klein et al. [12] measure "colocation" events, in which reported locations of two devices lie within a roughly 60 square foot spatial grid cell. Couture et al. [16] compute a "device exposure index" that measures the colocation of devices within a sample of preselected venues like restaurants or retail establishments. Chang et al. [30] use colocation matrices from Facebook [36] that measure the probability that devices from different geographic areas appear in the same 600-meter square region for five minutes, aggregated by week. Morley et al. [37] use the "human encounters" metric from Unacast [38, 39] that measures the frequency of two devices being within 50 meters of each other for an hour or less. Finally, Cuebiq offers a contact index measuring when two or more devices are within 50 feet of each other within five minutes [40].

Existing mobility metrics might not capture simultaneous colocation of devices, do not measure contact within a two-meter distance associated with highest transmission risk (via direct contact or exposure to respiratory

droplets), and might not take intrinsic mobile device spatial location error (horizontal uncertainty) into account. A better measure of contact events – the primary behavioral risk factor for transmission – could help explain historical patterns of transmission, assist policymakers in targeting interventions and messaging campaigns to encourage social distancing, guide public health response measures such as enhanced testing and contact tracing, and provide early warning to detect and prevent emerging outbreaks. By using highly detailed mobile device geolocation data and a novel probabilistic method for assessing close proximity, we sought to quantify total intensity of close interpersonal contact (within six feet) at the population-level (contact rate) and to use contact rate to explain patterns of COVID-19 incidence and predict emergence of new COVID-19 cases in the state of Connecticut, U.S. during February 1, 2020 – January 31, 2021.

Setting: Connecticut

Connecticut (population 3.565 million), like other states in the northeastern U.S., experienced a strong initial wave of COVID-19 infections during March–April 2020 following outbreaks in the New York City area [41, 42]. On March 17, Connecticut Governor Ned Lamont closed schools [43–46], and issued a statewide "Stay Safe, Stay Home" mandate to take effect on March 23, 2020 [47]. Governor Lamont's executive order recommended that nonessential businesses cease all in-person functions, closed in-person dining at restaurants, and cancelled all in-person community gatherings. The mandate excluded healthcare, food service, law enforcement, and other essential services.

As case counts declined, Connecticut followed a gradual reopening plan designed to resume economic activity while minimizing the risk of transmission via close contact between people. On May 20, the state entered Phase 1, permitting the following to open at 50% capacity with social distancing: hair salons and barbershops, outdoor zoos and museums, outdoor dining, outdoor recreation, retail shopping, university research, and offices, although work from home was strongly encouraged [48]. On June 17, Phase 2 began, permitting indoor religious services at 25% capacity and capped at 100 people, outdoor religious services capped at 150 people, and opening indoor dining, hair salons, personal service businesses, and libraries at 50% capacity [49]. A serology study to measure prevalence of SARS-CoV-2-specific IgG antibodies was conducted among adult Connecticut residents residing in non-congregate settings during June–July [50]. The study estimated a seroprevalence of 4.0% (90% confidence interval 2.0%–6.0%). Participants in the study reported their risk mitigation behaviors: 73% avoided public places, 75% avoided gatherings of families or friends, and 97% wore a mask at least some

of the time. In July, Governor Lamont delayed the state's planned summer move to Phase 3 – which would have loosened occupancy restrictions on bars and restaurants – because of surges in transmission occurring elsewhere in the U.S. [51].

Connecticut experienced low COVID-19 incidence and declining hospitalization during June–August, but in August a major outbreak occurred in Danbury, a town in the western part of the state [52]. During August–September, in-person education resumed at many colleges, universities, and primary/secondary schools in Connecticut. By mid-September, the state was facing a broad resurgence of COVID-19 transmission. On September 17, the Connecticut Department of Public Health reported that the number of new cases per week for the previous 4 weeks was 62% higher than the average number of new cases per week in July and early August [53]. These signs of resurgence were initially concentrated in southeastern Connecticut, where few COVID-19 cases were identified during the initial spring wave in March–April [54, 55].

Public health officials identified travel, social gatherings, workplaces, churches, universities, and recreational sports as contributing to transmission [56]. Nevertheless, on October 8, the state began Phase 3 reopening, permitting 50% capacity in houses of worship capped at 200 people, uncapped outdoor religious gatherings with social distancing, and opening indoor dining, hair salons, personal service businesses, and libraries at 75% capacity [49]. On November 6, as COVID-19 case counts continued to increase, Connecticut reverted to "Phase 2.1", reducing indoor restaurant seating, indoor and outdoor event capacity, and placing caps on attendance [49]. Case counts increased through December and began to decline in January 2021. Overall, Connecticut residents complied with state guidelines and mandates to reduce close contact. In a survey of risk mitigation behaviors throughout the U.S., Lazer et al. [7] reported that Connecticut ranked 9th among U.S. states in self-reported social distancing during fall 2020, and 6th in self-reported mask wearing. But case counts indicate that Connecticut experienced widely varying temporal and geographic dynamics of COVID-19 incidence over the course of the pandemic.

Computing contacts from anonymized mobile device geolocation data

We obtained anonymized mobile device geolocation data for a sample of devices in Connecticut from X-Mode. During May 1, 2020 through January 31, 2021, we observed a total of 788,842 unique (anonymized) device IDs, representing roughly 22% of the approximately 3.565 million residents of Connecticut (though some of those devices may have belonged to people residing elsewhere). An average of 141,617 unique devices were

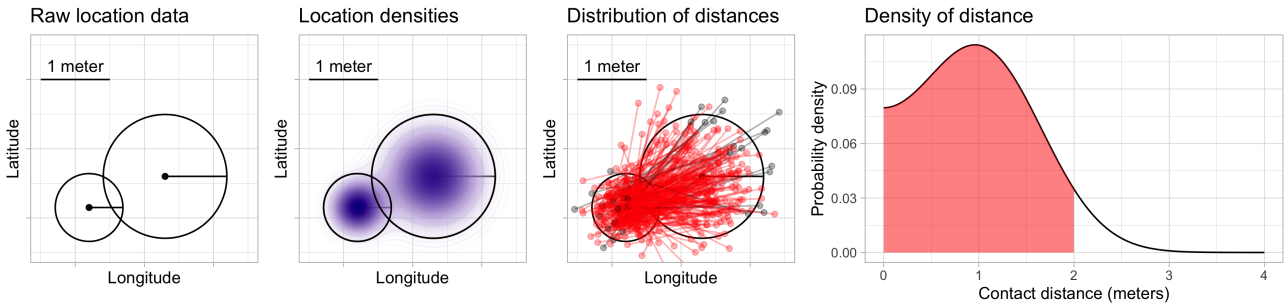

Figure 1: Schematic illustration of contact probability calculation. From left to right: raw locations, including horizontal uncertainty estimates, for two mobile devices are transformed into approximate location probability densities. The distribution of distances from points drawn randomly from these densities is computed. Sampled distances are shown here for illustrative purposes in red (when sampled device locations are within six feet apart) and gray (when sampled locations are more than six feet apart); in our implementation, the distribution of these distances is computed analytically. The shaded area under the density is the probability that the devices are within six feet.

observed per day. For each week, an average of 80.5% of device IDs from the prior week were present in the data. Devices might not be present in the dataset if the user turns off the device or does not interact with applications that report location data. Using device geolocation records consisting of anonymized device IDs, GPS coordinates, date/time stamps, and GPS location error estimates (horizontal uncertainty), we calculated the location in which each device had the most location records and designated that area as the device's primary dwell location (i.e., town of residence of device owner).

A contact event was computed by using a probabilistic algorithm that computes the likelihood of simultaneous 2-meter proximity between pairs of devices across geographic areas. For each device, we identify sets of records where devices were in spatial proximity to one another and stationary. A limitation of mobile device gelocation data is that it is not possible to precisely quantify the duration a device is stationary because device locations are collected asynchronously and irregularly over time. For each potential contact event, we compute the probability that the two device locations are within six feet by assuming that the reported device locations arise from a two-dimensional Gaussian probability distribution whose variance is computed by using the horizontal uncertainty measure, and correct the distance to account for the curvature of the earth. Figure 1 shows a schematic illustration of the contact event probability calculation.

We define the "contact rate" as the total number of contact events per day among observed devices at the town level; the contact rate is computed by summing daily contact probabilities for each device and assigning that sum to the device primary dwell location. A detailed description of the mobile device geolocation data, computation of the probability of contact, spatial aggregation of the contact probabilities to estimate contact rate, and coverage of mobile devices across Connecticut is given in the Supplementary Materials.

The mobile device geolocation and COVID-19 case data contain no individually identifying meta-data and were aggregated by day and town. This work was approved by the Yale University institutional review board. This work was also reviewed by CDC and was conducted consistently with applicable federal law and CDC policy [1].

Statewide contact trends

Figure 2 shows the contact rate by town in Connecticut during February 1 – January 31, 2021. Maps show the weekly average of daily contact rate by town, where darker colors in maps indicate higher contact rate. The daily contact rate is shown in the plot below. The statewide contact rate dropped dramatically in March, about one week before Governor Lamont issued the statewide stay-at-home mandate on March 23 [47]. News of surging COVID-19 hospitalization and responses in the New York area [57, 58], closure of public schools, and anticipation of a possible stay-at-home order might have played a role in reducing contact before the mandate was announced. After staying low during most of April, the contact rate began to rise slowly throughout the state during June–August. Incidence of infection was likely much higher during the first wave than the second, but steadily increasing availability of SARS-CoV-2 testing yielded higher case counts in the second wave. An interactive web application for exploring the contact rate in Connecticut is available at `https://forrestcrawford.shinyapps.io/ct_social_distancing`. The Supplementary Materials describe the web application in detail.

The Supplementary Material presents a comparison of the contact rate to mobility metrics from Google [59], Apple [60], Facebook [61], Descartes Labs [14, 62], and Cuebiq [40]. Most mobility metrics provided by these companies returned to values near the February/March baseline by the beginning of July. In contrast, the contact rate shown in Figure 2 shows that close interpersonal contact stayed low and rose slowly during June–August, 2020. Mobility metrics returned more quickly to the February 2020 baseline (or higher) compared to the contact rate and do not explain the low COVID-19 incidence achieved in Connecticut during June–August, 2020.

One explanation for the discrepancy between close contact and mobility metrics is that it is possible to travel far from home, to many distinct points of interest, or to many geographic areas, without coming into close contact with others. This might be what occurred in the summer of 2020: as Connecticut began its phased reopening plan, people resumed more normal patterns of away-from-home movement – work, shopping, or recreational activities – while maintaining social distancing. For this reason, when mobility metrics are used as

[1] See e.g., 45 C.F.R. part 46, 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq.

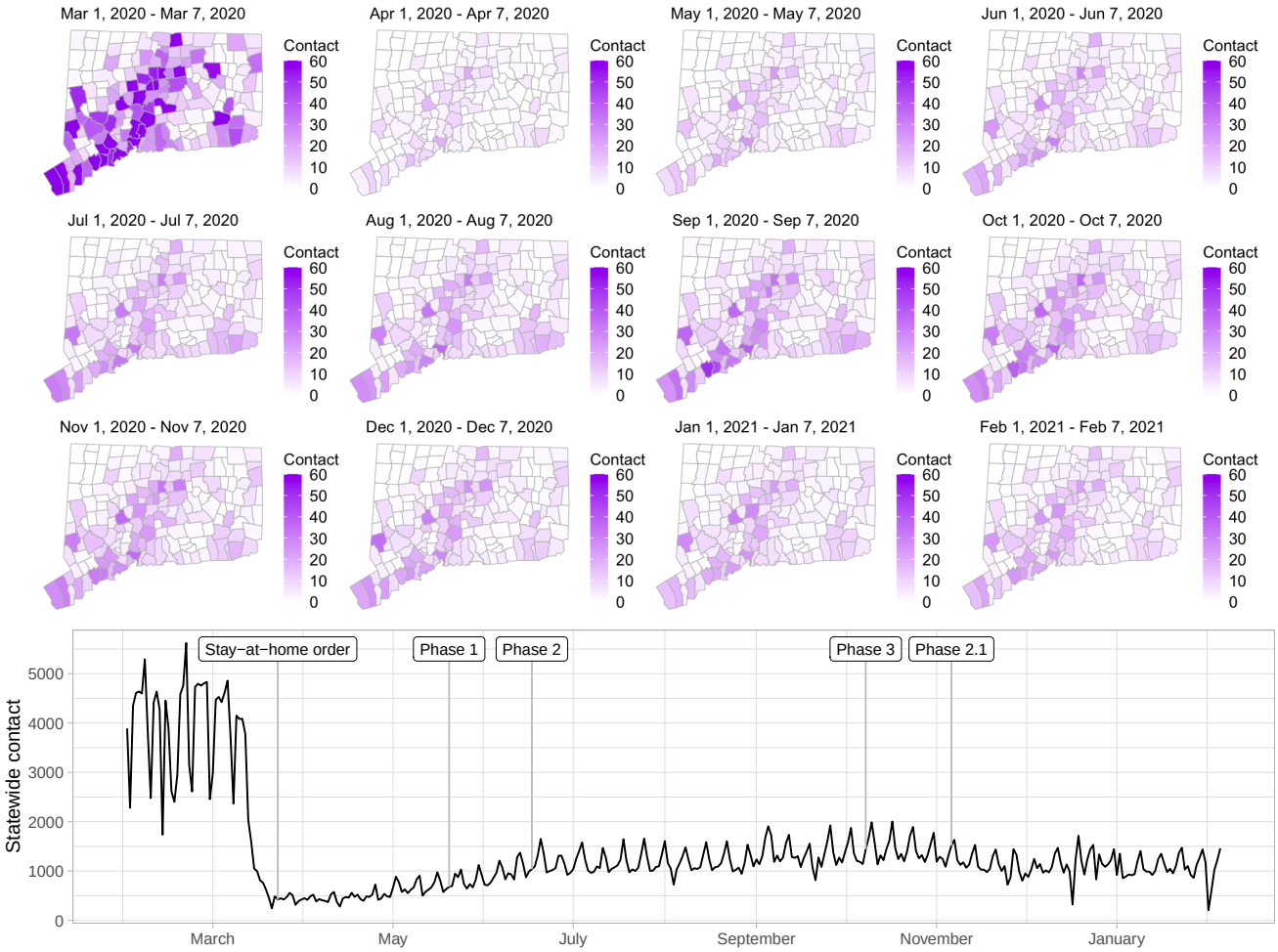

Figure 2: Estimated contact rate among mobile devices in our dataset in Connecticut from February 2020 to February 2021. At top, maps show the number of contacts in Connecticut's 169 towns per day during weeks beginning on the first of each month. Darker colors indicate higher contact. At bottom, statewide contact shows the daily frequency of close contact within six feet between distinct devices in our dataset. Governor Ned Lamont's stay-at-home order and reopening phases 1, 2, 3, and 2.1 indicated. The state reverted to the more restrictive "Phase 2.1" in response to rising case counts in November.

proxy measures of close interpersonal contact, they might overstate the risk of disease transmission.

Prediction of COVID-19 cases in Connecticut towns

To evaluate the contact rate as a predictor of COVID-19 burden in Connecticut, we use confirmed COVID-19 case data from non-congregate settings reported to the Connecticut Department of Public Health. We excluded cases among residents of long-term care facilities, managed residential communities (e.g., assisted living facilities), or correctional institutions. We aggregated non-congregate case data by day of sample collection, by town. We obtained town-level population estimates from the American Community Survey [63, 64].

We predict transmission of SARS-CoV-2 and COVID-19 cases in Connecticut towns using a continuous-time deterministic compartmental transmission model based on the susceptible-exposed-infective-removed (SEIR) process [65]. We accommodate geographical variation in transmission within Connecticut and estimated features of COVID-19 disease progression, hospitalization, and death. The model incorporates flexible time-varying case-finding rates at the town level. We incorporate the contact rate into the time-varying transmission risk by multiplying the standardized contact rate by the product of the baseline transmission rate and the estimated number of susceptible and infectious individuals in each town. We fit the model to statewide data, and produce model projections for each of Connecticut's 169 towns using the town population size, time-varying contact rate, estimated initial infection fraction, and time-varying case-finding rate. The model is conceptually similar to other SEIR-type COVID-19 transmission models making use of mobility data, but incorporates much geographic variation in transmission rates [26, 66–74]. The model and calibration procedure are described in detail in [65] and in the Supplementary Materials.

Figure 3 shows contact rates, estimated SARS-CoV-2 infections, observed and estimated case counts, estimated cumulative incidence, as well as 95% uncertainty intervals for model estimates, for the five largest cities by population in Connecticut: Bridgeport, Hartford, New Haven, Stamford, and Waterbury. Contact rates in these towns largely mirror rates in the state as a whole. Model estimates track the pattern of case counts through the full course of the epidemic, including the dramatic reduction in transmission during June–August. In some towns, e.g. Stamford, case counts are under-estimated in model projections during the first wave during March–April 2020. In these cases, dynamics of SARS-CoV-2 infections may differ from the dynamics of case counts because the estimated case detection rate (via viral testing) varied dramatically over time and geography.

Role of contact in local outbreaks

As COVID-19 case counts in Connecticut decreased during June–August, new and more heterogeneous patterns of transmission emerged. Figure 4 shows contact rates, confirmed non-congregate COVID-19 case counts, and 95% uncertainty intervals for cases in five Connecticut towns where incidence patterns differed from those of the larger cities shown in Figure 3.

During June–August, the only known community-wide COVID-19 outbreak in Connecticut occurred in the town of Danbury (population 84,479) [52]. During August 2–20, at least 178 new COVID-19 cases were reported,

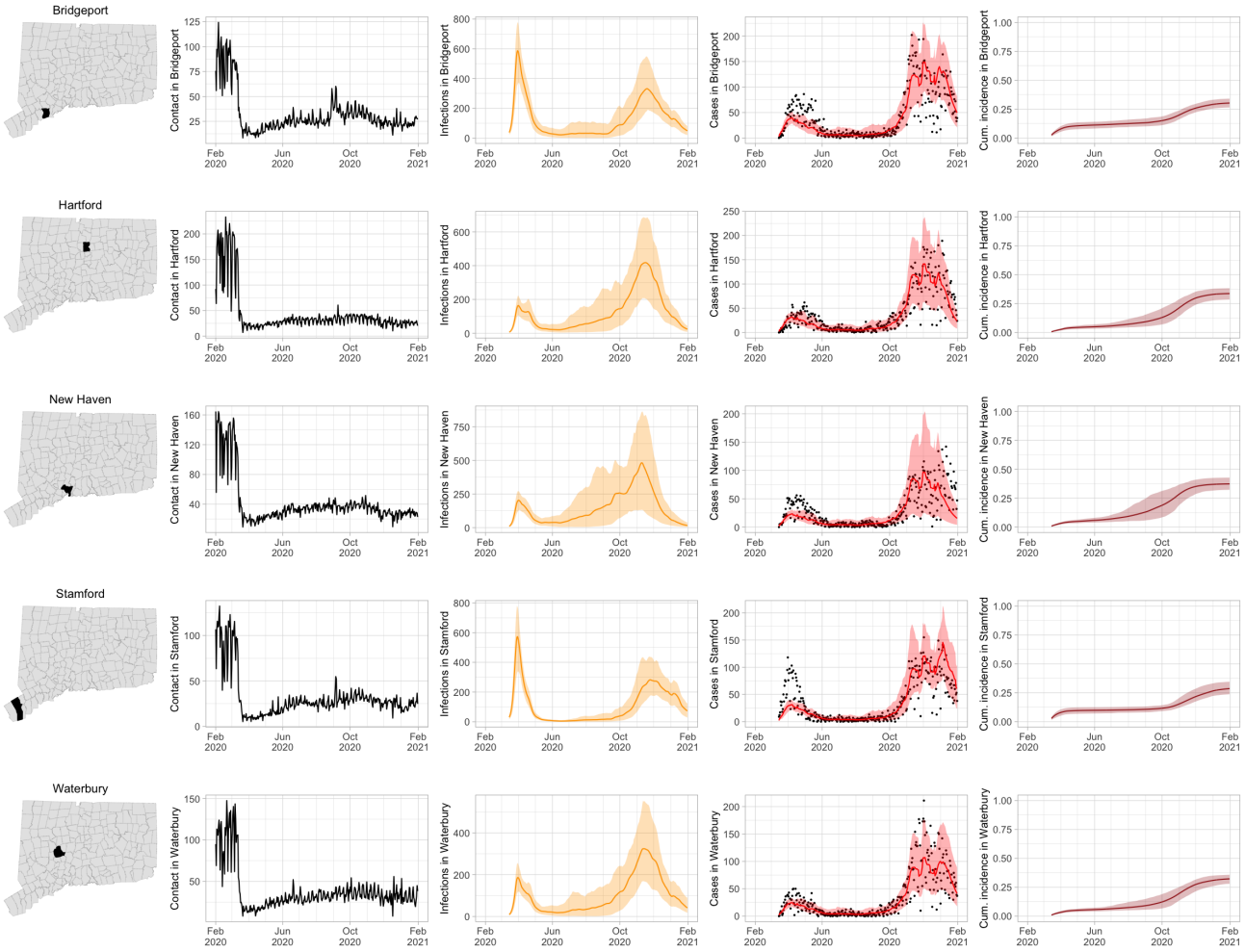

Figure 3: Contact rates, COVID-19 cases, and model predictions (with 95% uncertainty intervals) of infections, cases, and cumulative incidence proportion in the five largest cities by population in Connecticut: Bridgeport, New Haven, Hartford, Stamford, and Waterbury. Black dots show confirmed non-congregate COVID-19 case counts.

a significant increase from 40 cases reported during the prior week. Contact tracing investigations by public health officials attributed the outbreak to travel, but the contact rate was high in Danbury beginning in July and genomic analyses suggested the outbreak was closely linked to lineages already circulating in New York City and Connecticut [75, 76]. Predictions from the model including contact rates from Danbury suggest that this outbreak might have been part of a long-term increase in infections that began earlier in July and continued mostly unabated through November.

The town of Fairfield, bordering the larger city of Bridgeport, has a population of 62,105 people, and contains two universities, both of which reopened for in-person education in mid-August. The university communities experienced a surge in cases during September–October after students returned [77]. Students had access to frequent COVID-19 testing, and test coverage in this community was likely higher than in the general population, so infections among students might have been more likely to be reported to public health authorities. Contact rates in both Fairfield and the adjacent city of Bridgeport increased (Figures 3 and 4) during September shortly after students arrived on campus. The consequence of this increase in contact rate is evident in the rise in case counts for Fairfield two to three weeks later.

The eastern part of Connecticut was largely spared in the first wave of infections during March–April, but Norwich (population 39,136) and nearby towns experienced a strong surge in cases beginning in mid-September [54, 55]. Contact rose more quickly in these towns, compared to the western part of the state, following the beginning of Phase 1 in May 2020. Low testing coverage during the spring and summer of 2020, imported infections from neighboring Rhode Island, and lower compliance with social distancing measures might have played a role in outbreaks in the eastern part of the state.

Contact data do not explain all variations in confirmed non-congregate COVID-19 case counts. Though the model fits cases well overall in large cities, it can fail to capture variation in case counts in smaller cities where testing coverage is lower, or in settings where case-finding effort varied over the time. For example, high case counts corresponding to outbreak investigations involving extensive testing in Danbury during August, and Norwich during September/October, do not directly reflect changes in contact, and are not captured by the model projections.

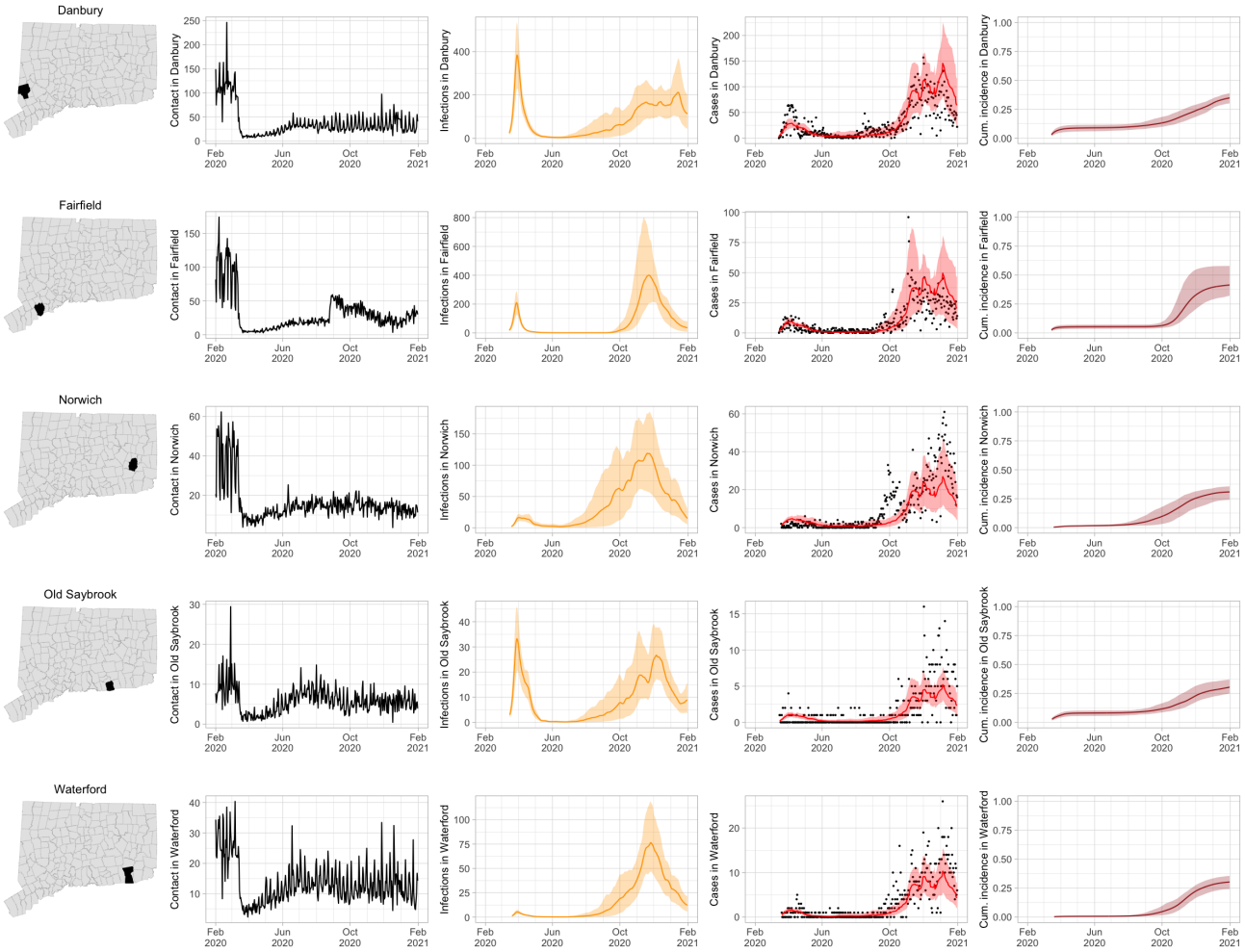

Figure 4: Contact rates, COVID-19 cases, and model predictions (with 95% uncertainty intervals) of infections, cases, and cumulative incidence proportion in several towns in Connecticut whose case or contact patterns differ from that of the state as a whole: Danbury, Fairfield, Norwich, Old Saybrook, and Waterford. Public health officials declared an outbreak in Danbury in mid-August 2020. Fairfield experienced outbreaks linked to two universities in September 2020. Norwich, Old Saybrook, and Waterford, in the eastern part of the state, were mostly spared during the first wave of infection, and had quickly rising case counts in fall 2020.

Contact may provide advance warning of COVID-19 cases

To assess the relationship between close interpersonal contact and COVID-19 cases without SEIR-type model assumptions about the dynamics of transmission, we fit a hierarchical Bayesian space-time statistical model to predict cases using town-level contact data. In a model that included 28 prior days of contact data, lagged contact from 3 to 7 days prior is significantly associated with current-day cases, in agreement with known features of the time to development of symptomatic disease [78–83]. An increase of 10 contacts in each of the previous 28 days within an average town gives rise to an increase in cases by a factor of 1.29 (95% credible interval [1.22, 1.37]) within that town. A model that includes contact predicts cases better than one without contact, according to goodness-of-fit criteria [84, 85]. The model structure and results are described in detail in the Supplementary Materials.

Discussion

Public health decision-makers track the COVID-19 pandemic using metrics – syndromic surveillance data, cases, hospitalizations, deaths – that lag disease transmission by days or weeks. In this paper, we have described a method for population-level surveillance of close interpersonal contact, the primary route for person-to-person transmission of SARS-CoV-2, by using anonymized mobile device geolocation data. The contact rate can reveal high-contact conditions likely to spawn local outbreaks, or areas where residents experience high contact rates, days or weeks before the resulting cases are detected by public health authorities through testing, traditional case investigation, and contact tracing. Because mobile device geolocation data are passively collected, contact rates are invariant to allocation and availability of public health resources for case finding. For this reason, contact rates could serve as a better early-warning signal for outbreaks than cases alone, especially when test volume is low. Contact rates could also have advantages over surveillance approaches using mobility metrics because interpersonal contact within six feet is more directly related to the likelihood of disease transmission by direct contact or respiratory droplets.

Contact rates could benefit public health efforts to prevent transmission of SARS-CoV-2 in two ways. First, community engagement programs could be directed to locations where the contact rate is high to improve social distancing practices or provide additional protective measures like ensuring adequate ventilation, environmental cleaning, and mask use. Second, enhanced testing in areas with high contact rates, and residential areas of

people experiencing that contact, could lead to earlier and more complete detection of cases. Earlier and more complete detection of cases enables faster and more complete isolation of cases and quarantine of contacts, which are crucial to stop transmission and stop outbreaks.

Contact rates also might be a useful addition to mathematical models of infectious disease transmission for prediction of COVID-19 infections or cases. In the early stages of the COVID-19 pandemic, researchers employed variations on the classical SEIR epidemic model [86, 87] to predict the initial wave of infections, estimate parameters like the basic reproduction number, and assess the effect of non-pharmaceutical interventions [e.g. 65, 88, 89, in Connecticut]. These models often assumed a constant population-level contact rate that is subsumed into a transmissibility parameter, or estimated contact rate from survey data collected prior to the pandemic [90, 91].

We have focused in this study on the U.S. state of Connecticut, but the usefulness of anonymized and passively collected contact data could be generalized to other settings. In the U.S., where mobile phone usage is high, states or towns can implement contact surveillance at low cost by working with private sector mobile device data providers. Like Connecticut, other states and countries experienced constrained testing availability in the early stages of the pandemic, and uneven geographic distribution of testing after test volume increased. Non-pharmaceutical interventions such as stay-at-home mandates, business and school closures, and social distancing guidelines also had uneven adoption and compliance varied across time and geography. Surveillance of contact rates could help officials better distribute testing resources and monitor intervention compliance in numerous settings. Internationally, mobile phone ownership has grown quickly but might be low in some developing countries [92], making contact surveillance less feasible in these settings.

The contact rate employed here has several advantages over existing mobility metrics and measures of mobile device density and proximity. First, the contact rate has been designed specifically to measure interpersonal contact within 6-feet relevant to COVID-19 transmission, as defined by CDC [1]. In contrast, mobility metrics primarily measure movement, which might not be a good proxy measure of close interpersonal contact. For each potential contact event between two devices, we use the reported device locations and horizontal uncertainty measurements to compute the probability that the devices were within six feet of one another. In this way, each potential contact event is weighted by the likelihood that the people carrying the devices were close enough for transmission to occur. In contrast, Unacast's "human encounters" metric measures the frequency of two devices being within 50 meters of one another. Because the Unacast definition includes interactions that are at a distance much farther than six feet, many are unlikely to involve the potential for disease transmission. The

contact rate used here incorporates close interpersonal contact occurring in every location in Connecticut, not only at pre-selected venues [e.g. 32, 93] therefore, the contact rate might be a better proxy for population-level transmission risk when there are prevalent infections.

Contact data derived from mobile device geolocation data have limitations. First, not all devices in Connecticut appear in the sample: during May 1– November 28, 2020, we observed a total of 788,842 unique device IDs, representing roughly 22% of the 3.6 million residents of Connecticut. An analysis in the Supplementary Materials shows that there is no evidence of systematic under-coverage of mobile devices as a function of town population sizes, but coverage declines slightly in towns with higher percent of residents identifying themselves as non-White, lacking a high school degree, and below the poverty level. Under-coverage among particularly vulnerable populations could result in under-counting of potential transmission events likely to affect these populations. Second, horizontal uncertainty varies by device and location, making close interpersonal contact that occurs in some areas more difficult to detect with certainty. Third, the duration of time a device was stationary is unknown because location data are reported asynchronously and at irregular intervals. Fourth, using anonymized mobile device geolocation data we do not observe individual-level demographic information, whether a potential contact occurred indoors or outdoors, nor additional individual-level infection risk factors or risk mitigation behavior like mask-wearing, hand washing, avoidance of touching surfaces or avoidance of crowded indoor spaces. However, CDC recommends the determination of close contact should be made "irrespective of whether the person with COVID-19 or the contact was wearing a mask" [94].

The contact rate might not detect all types of close interpersonal contact relevant for disease transmission and does not distinguish between physical contact and close proximity. We exclude contact occurring at primary dwell locations, so contact between pairs of people while at their shared same primary dwell locations is not represented in the contact rate. As a result our model projections may not adequately capture household transmission. Close contact that occurs while traveling, for example on a bus or train, might not be detected because devices are not stationary. Devices located on different floors of the same building might report nearby locations, even if the devices are separated by one or more floors. Location information might be reported by each device at irregular intervals, so we might not observe some kinds of fleeting contact. Contact that occurs outside of Connecticut is not recorded in our dataset. In particular, we did not observe information about contact for people who live in Connecticut and work in the New York City area.

Statewide contact rate based on mobile device geolocation data helps explain Connecticut's success in avoiding a broad resurgence in COVID-19 cases during June–August 2020, emergence of localized outbreaks during

late August–September, and a broad statewide resurgence during October–December. In addition to explaining historical patterns of transmission, incorporating contact rates into an SEIR transmission model might improve prediction of future COVID-19 cases and outbreaks at the town level, which could inform targeted allocation of public health prevention measures, such as SARS-CoV-2 testing and contact tracing with subsequent isolation or quarantine. Contact rate estimated from mobile device geolocation data can help improve population-level surveillance of close interpersonal contact, guide public health messaging campaigns to encourage social distancing, and in allocation of testing resources to detect or prevent emerging local outbreaks.

Acknowledgements

We are grateful to Gary Archambault, Anderson Brito, Maciej F Boni, Deidre Gifford, Gregg Gonsalves, Nathan Grubaugh, Edward Kaplan, Albert Ko, Jessica Lee, Sean Nolan, Saad Omer, Jasjeet Sekhon, Olivia Schultes, Mary Souther, Lynn Sosa, Robert Tollefson, and Robert Zalot. We thank Stony Brook Research Computing and Cyberinfrastructure, and the Institute for Advanced Computational Science at Stony Brook University. We thank X-Mode Social for providing the source data used to develop the contact metrics. We are grateful fo aggregated mobility data provided by Google, Apple, Facebook, Descartes Labs, and Cuebiq. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

Funding: This work was supported by NIH grants NICHD 1DP2HD091799-01, NIAID R01AI137093-03, Cooperative Agreement 6NU50CK000524-01 from the Centers for Disease Control and Prevention, funds from the COVID-19 Paycheck Protection Program and Health Care Enhancement Act, and the Pershing Square Foundation. Computing resources at Stony Brook University were funded by National Science Foundation grant # 1531492.

Authors contributions: Conceptualization: FWC, JB, SJ; Methodology Development: FWC, OM, ZRL, JLW, JB, JC, PK, TV; Software Development: FWC, OM, ZRL, SJ, JLW, SGD, JC, PK, TV; Validation: FWC, OM, SGD; Formal Analysis: FWC, OM, JLW; Resources: FWC, JB, MC; Data Curation: SJ, OM, JB, JC, PK, TV; Writing – Original Draft: FWC; Writing – Review & Editing: FWC, SJ, OM, ZRL, SGD, JLW, TV, JC, JB, MC; Visualization Preparation: FWC, SGD, OM, ZRL; Supervision Oversight: FWC, JB, MC; Project Administration: FWC, JB; Funding Acquisition: FWC, JB, MC.

Competing interests: FWC is a paid consultant to Whitespace Solutions, Ltd. JB, JC, PK, and TV are employ-

ees of Whitespace Solutions, Ltd.

List of Supplementary Materials

Materials and Methods

Supplementary Text

Table S1–S3

Fig S1–S32

References

References

[1] Centers for Disease Control and Prevention. Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission. https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html, Updated Oct. 5, 2020.

[2] Centers for Disease Control and Prevention. Social Distancing. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html, Updated Nov. 17, 2020.

[3] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, Mélodie Monod, Imperial College COVID-19 Response Team, Azra C. Ghani, Christl A. Donnelly, Steven Riley, Michaela A. C. Vollmer, Neil M. Ferguson, Lucy C. Okell, and Samir Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020. doi: https://doi.org/10.1038/s41586-020-2405-7.

[4] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy C Okell, Sabine van Elsland, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick GT Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A Donnelly, Steven Riley, and Azra C Ghani. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. 2020. doi: https://doi.org/10.25561/77482.

[5] Shengjie Lai, Nick W Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R Floyd, Amy

Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*, 585:410–413, 2020. doi: https://doi.org/10.1038/s41586-020-2293-x.

[6] Abiel Sebhatu, Karl Wennberg, Stefan Arora-Jonsson, and Staffan I Lindberg. Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. *Proceedings of the National Academy of Sciences*, 117(35):21201–21208, 2020. doi: https://doi.org/10.1073/pnas.2010625117.

[7] David Lazer, Mauricio Santillana, Roy H. Perlis, Alexi Quintana, Katherine Ognyanova, Jonathan Green, Matthew A. Baum, Matthew Simonson, Ata A. Uslu, Hanyu Chwe, James Druckman, Jennifer Lin, and Adina Gitomer. The COVID states project: A 50-state COVID-19 survey report 26: Trajectory of COVID-19-related behaviors. *The COVID-19 Consortium for Understanding the Public's Policy Preferences Across States*, 2020. URL https://kateto.net/covid19/COVID19%20CONSORTIUM%20REPORT%2026%20BEHAVIOR%20NOV%202020.

[8] Caroline O Buckee, Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, Yonatan H Grad, Bryan Grenfell, M Elizabeth Halloran, Moritz UG Kraemer, Marc Lipsitch, C. Jessica E. Metcalf, Lauren Ancel Meyers, T. Alex Perkins, Mauricio Santillana, Samuel V. Scarpino, Cecile Viboud, Amy Wesolowski, and Andrew Schroeder. Aggregated mobility data could help fight COVID-19. *Science*, 368(6487):145, 2020.

[9] Kyra H Grantz, Hannah R Meredith, Derek AT Cummings, C Jessica E Metcalf, Bryan T Grenfell, John R Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, Caroline O. Buckee, and Amy Wesolowski. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications*, 11(4961), 2020.

[10] Nuria Oliver, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sébastien Deletaille, Marco De Nadai, Emmanuel Letouzé, Albert Ali Salah, Richard Benjamins, Ciro Cattuto, Vittoria Colizza, Nicolas de Cordes, Samuel P. Fraiberger, Till Koebe, Sune Lehmann, Juan Murillo, Alex Pentland, Phuong N Pham, Frédéric Pivetta, Jari Saram aki, Samuel V. Scarpino, Michele Tizzoni, Stefaan Verhulst, and Patrick Vinck. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6(23):eabc0764, 2020. doi: 10.1126/sciadv.abc0764.

[11] Stephen Kissler, Nishant Kishore, Malavika Prabhu, Dena Goffman, Yaakov Beilin, Ruth Landau, Cynthia Gyamfi-Bannerman, Brian Bateman, Daniel Katz, Jonathan Gal, Angela Bianco, Joanne Stone, Daniel Larremore, Caroline Buckee, and Yonatan Grad. Reductions in commuting mobility predict geographic

differences in SARS-CoV-2 prevalence in New York City. *Nature Communications*, 11(4674), 2020. doi: 10.1038/s41467-020-18271-5.

[12] Brennan Klein, Timothy LaRock, Stefan McCabe, Leo Torres, Lisa Friedland, Filippo Privitera, Brennan Lake, Moritz U. G. Kraemer, John S. Brownstein, David Lazer, Tina Eliassi-Rad, Samuel V. Scarpino, Alessandro Vespignani, and Matteo Chinazzi. Reshaping a nation: Mobility, commuting, and contact patterns during the COVID-19 outbreak. 2020. URL `https://www.mobs-lab.org/uploads/6/7/8/7/6787877/covid19mobility_report2.pdf`.

[13] Martin Andersen. Early evidence on social distancing in response to COVID-19 in the United States. *SSRN*, April 2020.

[14] Michael S Warren and Samuel W Skillman. Mobility changes in response to COVID-19. *arXiv preprint arXiv:2003.14228*, March 2020. URL `https://arxiv.org/abs/2003.14228`.

[15] Brennan Klein, Timothy LaRock, Stefan McCabe, Leo Torres, Filippo Privitera, Brennan Lake, Moritz U. G. Kraemer, John S. Brownstein, David Lazer, Tina Eliassi-Rad, Samuel V. Scarpino, Matteo Chinazzi, and Alessandro Vespignani. Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak. 2020. URL `https://www.mobs-lab.org/uploads/6/7/8/7/6787877/assessing_mobility_changes_in_the_united_states_during_the_covid_19_outbreak.pdf`.

[16] Victor Couture, Jonathan I Dingel, Allison E Green, Jessie Handbury, and Kevin R Williams. Measuring movement and social contact with smartphone data: A real-time application to COVID-19. Technical report, National Bureau of Economic Research, 2020.

[17] Joshua Coven and Arpit Gupta. Disparities in mobility responses to COVID-19. Technical report, NYU Stern Working Paper, 2020.

[18] Milena Almagro and Angelo Orane Hutchinson. The Determinants of the Differential Exposure to COVID-19 in New York City and Their Evolution Over Time. *SSRN*, June 2020.

[19] Shengjie Lai, Isaac I Bogoch, Nick W Ruktanonchai, Alexander Watts, Xin Lu, Weizhong Yang, Hongjie Yu, Kamran Khan, and Andrew J Tatem. Assessing spread risk of Wuhan novel coronavirus within and beyond China, January-April 2020: A travel network-based modelling study. *medRxiv*, 2020. doi: https://doi.org/10.1101/2020.02.04.20020479.

[20] Emanuele Pepe, Paolo Bajardi, Laetitia Gauvin, Filippo Privitera, Brennan Lake, Ciro Cattuto, and Michele Tizzoni. COVID-19 outbreak response: a first assessment of mobility changes in Italy following national lockdown. *Scientific Data*, 7(230), 2020. doi: https://doi.org/10.1038/s41597-020-00575-2.

[21] Matthew Abueg, Robert Hinch, Neo Wu, Luyang Liu, William JM Probert, Austin Wu, Paul Eastham, Yusef Shafi, Matt Rosencrantz, Michael Dikovsky, et al. Modeling the combined effect of digital exposure notification and non-pharmaceutical interventions on the COVID-19 epidemic in Washington state. *medRxiv*, 2020.

[22] Corentin Cot, Giacomo Cacciapaglia, and Francesco Sannino. Mining Google and Apple mobility data: temporal anatomy for COVID-19 social distancing. *Scientific Reports*, 11(1):1–8, 2021.

[23] Hamada S Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M Squire, and Lauren M Gardner. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254, July 2020.

[24] Samuel Engle, John Stromme, and Anson Zhou. Staying at home: Mobility effects of COVID-19. *SSRN*, April 2020. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3565703`.

[25] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, Marco Ajelli, and Hongjie Yu. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, 368(6498):1481–1486, 2020.

[26] Andrew C Miller, Nicholas J Foti, Joseph A Lewnard, Nicholas P Jewell, Carlos Guestrin, and Emily B Fox. Mobility trends provide a leading indicator of changes in SARS-CoV-2 transmission. *medRxiv*, 2020.

[27] Parker Liautaud, Peter Huybers, and Mauricio Santillana. Fever and mobility data indicate social distancing has reduced incidence of communicable disease in the United States. *arXiv preprint arXiv:2004.09911*, April 2020. URL `https://arxiv.org/abs/2004.09911`.

[28] Armando Cartenì, Luigi Di Francesco, and Maria Martino. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Science of the Total Environment*, 741:140489, 2020. doi: https://doi.org/10.1016/j.scitotenv.2020.140489.

[29] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Open COVID-19 Data Working Group, Louis Du Plessis, Nuno R Faria, Ruoran Li, William P Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490):493–497, 2020. doi: 10.1126/science.abb4218.

[30] Meng-Chun Chang, Rebecca Kahn, Yu-An Li, Cheng-Sheng Lee, Caroline O Buckee, and Hsiao-Han Chang. Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan. *medRxiv*, 2020. doi: https://doi.org/10.1101/2020.04.07.20053439.

[31] David Martín-Calvo, Alberto Aleta, Alex Pentland, Yamir Moreno, and Esteban Moro. Effective-

ness of social distancing strategies for protecting a community from a pandemic with a data driven contact network based on census and real-world mobility data. *MIT Connection Science Preprint*, 2020. URL https://connection.mit.edu/sites/default/files/publication-pdfs/Preliminary_Report_Effectiveness_of_social_distance_strategies_COVID-19%20%281%29.pdf.

[32] Serina Y Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 2020.

[33] Jonathan Jay, Jacob Bor, Elaine O Nsoesie, Sarah K Lipson, David K Jones, Sandro Galea, and Julia Raifman. Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. *Nature Human Behaviour*, pages 1–9, 2020.

[34] Luyang Liu, Sharad Vikram, Junpeng Lao, Xue Ben, Alexander D'Amour, Shawn O'Banion, Mark Sandler, Rif A Saurous, and Matthew D Hoffman. Estimating the changing infection rate of COVID-19 using Bayesian models of mobility. *medRxiv*, 2020.

[35] Centers for Disease Control and Prevention. Explore human mobility and COVID-19 transmission in your local area, 2020. URL https://covid.cdc.gov/covid-data-tracker/#mobility.

[36] Facebook. Data for Good: New Tools to Help Health Researchers Track and Combat COVID-19. https://about.fb.com/news/2020/04/data-for-good/, 2020.

[37] Christopher P Morley, Kathryn B Anderson, Jana Shaw, Telisa Stewart, Stephen J Thomas, and Dongliang Wang. Social Distancing Metrics and Estimates of SARS-CoV-2 Transmission Rates: Associations Between Mobile Telephone Data Tracking and R. *Journal of Public Health Management and Practice*, 26(6): 606–612, 2020. doi: 10.1097/PHH.0000000000001240.

[38] Madeline Ngo. Rounding out the Social Distancing Scoreboard. https://www.unacast.com/post/rounding-out-the-social-distancing-scoreboard, 2020.

[39] Unacast. Social Distancing Scoreboard. https://www.unacast.com/covid19/social-distancing-scoreboard, 2020.

[40] Cuebiq. Mobility Insights. https://www.cuebiq.com/visitation-insights-covid19/, 2020.

[41] Jordan Peccia, Alessandro Zulli, Doug E Brackney, Nathan D Grubaugh, Edward H Kaplan, Arnau Casanovas-Massana, Albert I Ko, Amyn A Malik, Dennis Wang, Mike Wang, et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology*, 38(10): 1164–1167, 2020.

[42] Edward H Kaplan, Dennis Wang, Mike Wang, Amyn A Malik, Alessandro Zulli, and Jordan Peccia. Aligning SARS-CoV-2 indicators via an epidemic model: application to hospital admissions and RNA detection in

sewage sludge. *Health Care Management Science*, pages 1–10, 2020.

[43] Ned Lamont. Executive Order No. 7C: Protection of public health and safety during COVID-19 pandemic and response – further suspension or modification of statutes, March 2020. URL https://portal.ct.gov/-/media/Office-of-the-Governor/Executive-Orders/Lamont-Executive-Orders/Executive-Order-No-7C.pdf.

[44] Ned Lamont. Executive Order No. 7L: Protection of public health and safety during COVID-19 pandemic and response - extension of school cancellation, municipal retiree reemployment, open fishing season and additional public health measures, March 2020. URL https://portal.ct.gov/-/media/Office-of-the-Governor/Executive-Orders/Lamont-Executive-Orders/Executive-Order-No-7L.pdf.

[45] Ned Lamont. Executive Order No. 7X: Protection of public health and safety during COVID-19 pandemic and response – renter protections, extended class cancellation and other safety measures, educator certification, food trucks for truckers, April 2020. URL https://portal.ct.gov/-/media/Office-of-the-Governor/Executive-Orders/Lamont-Executive-Orders/Executive-Order-No-7X.pdf.

[46] Ned Lamont. Executive Order No 7II: Protection of public health and safety during COVID-19 pandemic and response – extension of school cancellation, home health care coverage, and food assistance measures, May 2020. URL https://portal.ct.gov/-/media/Office-of-the-Governor/Executive-Orders/Lamont-Executive-Orders/Executive-Order-No-7II.pdf.

[47] Ned Lamont. Executive Order No. 7H: Protection of public health and safety during COVID-19 pandemic and response - restrictions on workplaces for non-essential businesses, coordinated response effort, March 2020. URL https://portal.ct.gov/-/media/Office-of-the-Governor/Executive-Orders/Lamont-Executive-Orders/Executive-Order-No-7H.pdf.

[48] Ned Lamont. Governor lamont releases rules for businesses under first phase of connecticut's reopening plans amid COVID-19. State of Connecticut Press Release, May 2020. URL https://portal.ct.gov/Office-of-the-Governor/News/Press-Releases/2020/03-2020/Governor-Lamont-Announces-First-Positive-Case-of-Novel-Coronavirus-Involving-a-Connecticut-Res

[49] Connecticut Department of Economic and Community Development. Sector rules and certification for reopen. https://portal.ct.gov/DECD/Content/Coronavirus-Business-Recovery/Sector-Rules-and-Certification-for-Reopen, 2020. URL https://portal.ct.gov/DECD/Content/Coronavirus-Business-Recovery/Sector-Rules-and-Certification-for-Reopen.

[50] Shiwani Mahajan, Rajesh Srinivasan, Carrie A Redlich, Sara K Huston, Kelly M Anastasio, Lisa Cashman, Dorothy S Massey, Andrew Dugan, Dan Witters, Jenny Marlar, Shu-Xia Li, Zhenqiu Lin, Domonique Hodge, Manas Chattopadhyay, Mark D Adams, Charles Lee, Lokinendi V Rao, Chris Stewart, Karthik Kuppusamy, Albert I Ko, and Harlan M Krumholz. Seroprevalence of SARS-CoV-2-Specific IgG antibodies among adults living in Connecticut: Post-infection prevalence (PIP) study. *American Journal of Medicine*, S0002-9343(20):30909–8, 2020. doi: 10.1016/j.amjmed.2020.09.024.

[51] Keith M. Phaneuf. Lamont: Full reopening of bars in mid-July must 'take a pause'. *CT Mirror*, 2020. URL https://ctmirror.org/2020/07/02/lamont-full-reopening-of-bars-in-mid-july-must-take-a-pause/.

[52] Connecticut Department of Public Health. DPH issues COVID-19 alert to Danbury community after significant spike in new cases, 2020. URL https://portal.ct.gov/DPH/Press-Room/Press-Releases---2020/DPH-Issues-COVID-19-Alert-to-Danbury-Community-After-Significant-Spike-in-New-Cases.

[53] Connecticut Department of Public Health. Covid-19 update september 17, 2020, 2020. URL https://portal.ct.gov/-/media/Coronavirus/CTDPHCOVID19summary9172020.pdf.

[54] Connecticut Department of Public Health. DPH issues COVID-19 alert for New London after significant rise in new cases, 2020. URL https://portal.ct.gov/DPH/Press-Room/Press-Releases---2020/DPH-ISSUES-COVID-19-ALERT-FOR-NEW-LONDON--AFTER-SIGNIFICANT-RISE-IN-NEW-CASES.

[55] Connecticut Department of Public Health. DPH issues COVID-19 alert for Norwich after significant spike in new cases, 2020. URL https://portal.ct.gov/DPH/Press-Room/Press-Releases---2020/DPH-Issues-COVID-19-Alert-for-Norwich-After-Significant-Spike-in-New-Cases.

[56] Alex Putterman. After a quiet summer, Connecticut's COVID-19 numbers are climbing again. Here's why. *Hartford Courant*, 2020. URL https://www.courant.com/coronavirus/hc-news-coronavirus-lamont-schools-covid-20200924-e2k66eyqnrgwxlwztn2ji5bgra-story.html.

[57] Corinne N Thompson, Jennifer Baumgartner, Carolina Pichardo, Brian Toro, Lan Li, Robert Arciuolo, Pui Ying Chan, Judy Chen, Gretchen Culp, Alexander Davidson, et al. COVID-19 outbreak–New York City, February 29–june 1, 2020. *Morbidity and Mortality Weekly Report*, 69(46):1725, 2020.

[58] Bill Chapell. Coronavirus: New York creates 'containment area' around cluster in New Rochelle, 2020. URL https://www.npr.org/sections/health-shots/2020/03/10/814099444/new-york-creates-containment-area-around-cluster-in-new-rochelle.

[59] Google. COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility, 2020.

[60] Apple. Mobility Trends Reports. `https://covid19.apple.com/mobility`, 2020.

[61] Facebook. COVID-19 Interactive Map & Dashboard. https://covid-survey.dataforgood.fb.com/?region=WORLD, 2020.

[62] Descartes Labs. Changes in Mobility. https://www.descarteslabs.com/mobility/, 2020.

[63] United States Census Bureau. *American Community Survey (ACS)*, 2020. URL `https://www.census.gov/programs-surveys/acs`.

[64] Connecticut Data Collaborative. *Data By Topic*, 2020. URL `http://data.ctdata.org/data_by_topic#demographics`.

[65] Olga Morozova, Zehang Richard Li, and Forrest W Crawford. A model for COVID-19 transmission in Connecticut. *medRxiv*, 2021. doi: https://doi.org/10.1101/2020.06.12.20126391. URL `https://www.medrxiv.org/content/10.1101/2020.06.12.20126391v2`.

[66] Centers for Disease Control and Prevention. COVID-19 Forecasts. `https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html`, 2020. Accessed: 2020-05-17.

[67] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, 2020.

[68] Stephen M Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H Grad, and Marc Lipsitch. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493):860–868, 2020.

[69] Tobias Brett and Pejman Rohani. COVID-19 herd immunity strategies: walking an elusive and dangerous tightrope. *medRxiv*, 2020.

[70] Francisco Perez-Reche and Norval Strachan. Importance of untested infectious individuals for the suppression of COVID-19 epidemics. *MedRxiv*, 2020. doi: https://doi.org/10.1101/2020.04.13.20064022.

[71] Marissa L Childs, Morgan P Kain, Devin Kirk, Mallory Harris, Lisa Couper, Nicole Nova, Isabel Delwel, Jacob Ritchie, and Erin A Mordecai. The impact of long-term non-pharmaceutical interventions on COVID-19 epidemic dynamics and control. *medRxiv*, 2020.

[72] Henrik Salje, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, Yann Le Strat, Justin Lessler, Daniel Levy Bruhl, Arnaud Fontanet, Lulla Opatowski, Pierre-Yves Boelle, and Simon Cauchemez. Estimating the burden of SARS-CoV-2 in France. *Science*, 369(6500):208–211, 2020.

[73] Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini Jr, Stefano Merler, Alex Pentland, Alessandro Vespignani, Esteban Moro, and Yamir Moreno. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour*, 4(9):964–971, 2020.

[74] COVID-19 Statistics, Policy modeling, and Epidemiology Collective and Joshua A Salomon. Defining high-value information for COVID-19 decision-making. *medRxiv*, 2020.

[75] Tara Alpert, Anderson Brito, and Nathan Grubaugh. Update 13: COVID Tracker CT, Yale SARS-CoV-2 Genomic Surveillance Initiative: Recent outbreaks in Danbury, CT, 2020. URL `https://covidtrackerct.com/portfolio/update13/`.

[76] Joseph Fauver, Mary Petrone, Tara Alpert, Anderson Brito, and Nathan Grubaugh. Update 12: Recent outbreaks in Danbury, CT, 2020. URL `https://covidtrackerct.com/portfolio/update12/`.

[77] Fairfield Department of Health. COVID-19 data page, 2020. URL `https://www.fairfieldct.org/covid19data`.

[78] Conor McAloon, Áine Collins, Kevin Hunt, Ann Barber, Andrew W Byrne, Francis Butler, Miriam Casey, John Griffin, Elizabeth Lane, David McEvoy, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ open*, 10(8):e039652, 2020.

[79] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 172 (9):577–582, 2020.

[80] Jantien A Backer, Don Klinkenberg, and Jacco Wallinga. Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, china, 20–28 january 2020. *Eurosurveillance*, 25(5), 2020.

[81] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy S.M. Leung, Eric H.Y. Lau, Jessica Y. Wong, Xuesen Xing, Nijuan Xiang, Yang Wu, Chao Li, Qi Chen, Dan Li, Tian Liu, Jing Zhao, Man Liu, Wenxiao Tu, Chuding Chen, Lianmei Jin, Rui Yang, Qi Wang, Suhua Zhou, Rui Wang, Hui Liu, Yinbo Luo, Yuan Liu, Ge Shao, Huan Li, Zhongfa Tao, Yang Yang, Zhiqiang Deng, Boxi Liu, Zhitao Ma, Yanping Zhang, Guoqing Shi, Tommy T.Y. Lam, Joseph T. Wu, George F. Gao, Benjamin J. Cowling, Bo Yang, Gabriel M. Leung, and Zijian Feng. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, 2020.

[82] Jing Qin, Chong You, Qiushi Lin, Taojun Hu, Shicheng Yu, and Xiao-Hua Zhou. Estimation of incubation

period distribution of COVID-19 using disease onset forward time: A novel cross-sectional and forward follow-up study. *Science Advances*, 6(33), 2020. doi: 10.1126/sciadv.abc1202. URL https://advances.sciencemag.org/content/6/33/eabc1202.

[83] Centers for Disease Control and Prevention. COVID-19 Pandemic Planning Scenarios. https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html, 2021. Accessed: 2021-01-18.

[84] Purushottam W Laud and Joseph G Ibrahim. Predictive model selection. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):247–262, 1995.

[85] Sumio Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594, 2010.

[86] Roy M Anderson, B Anderson, and Robert M May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1992.

[87] Matt J Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2011.

[88] Forrest W Crawford, Zehang Richard Li, and Olga Morozova. COVID-19 projections for reopening Connecticut. *medRxiv*, 2020. doi: https://doi.org/10.1101/2020.06.16.20126425. URL https://www.medrxiv.org/content/10.1101/2020.06.16.20126425v1.

[89] Melanie H Chitwood, Marcus Russi, Kenneth Gunasekera, Joshua Havumaki, Virginia E Pitzer, Joshua L Warren, Daniel Weinberger, Ted Cohen, and Nicolas A Menzies. Bayesian nowcasting with adjustment for delayed and incomplete reporting to estimate COVID-19 infections in the United States. *medRxiv*, 2020. doi: https://doi.org/10.1101/2020.06.17.20133983.

[90] Kiesha Prem, Alex R Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Computational Biology*, 13(9):e1005697, 2017. doi: https://doi.org/10.1371/journal.pcbi.1005697.

[91] Kiesha Prem, Kevin van Zandvoort, Petra Klepac, Rosalind M Eggo, Nicholas G Davies, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Alex R Cook, and Mark Jit. Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era. *medRxiv*, 2020. doi: https://doi.org/10.1101/2020.07.22.20159772.

[92] Kyle Taylor and Laura Silver. Smartphone ownership is growing rapidly around the world, but not always equally. *Pew Research Center*, 5, 2019.

[93] Mélodie Monod, Alexandra Blenkinsop, Xiaoyue Xi, Daniel Hebert, Sivan Bershan, Simon Tietze, Marc Baguelin, Valerie C Bradley, Yu Chen, Helen Coupland, et al. Age groups that sustain resurging COVID-

19 epidemics in the United States. *Science*, 2021.

[94] Centers for Disease Control and Prevention. Public Health Guidance for Community-Related Exposure. https://www.cdc.gov/coronavirus/2019-ncov/php/public-health-recommendations.html, Updated Dec. 3, 2020.

Supplement to "Impact of close interpersonal contact on COVID-19 incidence: evidence from one year of mobile device data"

Forrest W. Crawford[1,2,3,4], Sydney A. Jones[5,6], Matthew Cartter[6,7], Samantha G. Dean[1], Joshua L. Warren[1], Zehang Richard Li[7], Jacqueline Barbieri[8], Jared Campbell[8], Patrick Kenney[8], Thomas Valleau[8], Olga Morozova[9]

1. Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA.
2. Department of Statistics & Data Science, Yale University, New Haven, CT, USA.
3. Department of Ecology & Evolutionary Biology, Yale University, New Haven, CT, USA.
4. School of Management, Yale University, New Haven, CT, USA.
5. Epidemic Intelligence Service, Centers for Disease Control & Prevention, Atlanta, GA, USA.
6. Infectious Diseases Section, Connecticut Department of Public Health, New Haven, CT, USA.
7. Department of Statistics, University of California, Santa Cruz, Santa Cruz, CA, USA.
8. Whitespace Solutions, Ltd, Alexandria, VA, USA.
9. Program in Public Health and Department of Family, Population and Preventive Medicine, Stony Brook University, NY, USA.

March 10, 2021

1 Materials & Methods

1.1 Overview: computing the contact rate

Anonymized mobile device geolocation data for devices located within the state of Connecticut are received from a third-party provider, X-Mode. Mobile device users consent to sharing of their location information when they accept the terms and conditions of the applications that report location data. Mobile devices also permit users to turn off location sharing. The mobile device geolocation records contain anonymized unique device IDs that persist over time, GPS coordinates, date/time stamps, and GPS location error estimates (also called horizontal uncertainty) measured in meters. The data used in the analysis presented in this paper contain no individually identifying meta-data.

From these records we calculate the location in which each device had the most location records and designate that area as the device's primary dwell location. The primary dwell location is used to determine whether distinct devices reside at the same location. A device's primary dwell location is considered to be the best estimate of

the home location of the device user. We do not use time of day to determine primary dwell location, as this could lead to mis-characterization of primary dwell location for people who work at night instead of during the daytime, for example.

For each device, we identify clusters of records as locations where other devices were stationary and nearby. When two devices were stationary and in proximity to one another at the same time, the device locations and their corresponding GPS location errors are used to calculate a probability that the devices were in close contact, within six feet (approximately two meters). The resulting information is used to generate a contact event record including the date, time, location and horizontal accuracy of each device involved in the contact, both device IDs, and the computed probability of contact between the devices.

To avoid measuring spurious contact between people who are not actually close to one another, or contact between people who live together, we do not record contacts that occur in some places. A buffered polygon derived from roadway center lines is used to determine if a given contact event occurred on a roadway. If so, then the contact record excluded from the calculation of the contact rate within that region. This could result in missing close contact that occurs in vehicles, such as on buses, trains, or carpooling. Similarly, all contact events for devices at their estimated primary dwell location are tagged and excluded when computing contact rates.

1.2 Measuring close contact

We consider a simplified version of geospatial location information on Euclidean plane $\mathbb{R}^2$. We modify this setting below to account for the curvature of the earth. Suppose that for device location point i we observe the triple (X_i, Y_i, R_i) where (X_i, Y_i) is the reported location (in longitude and latitude) of the mobile device and R_i is the radius of horizontal uncertainty associated with the device location. We assume that the horizontal uncertainty radius R_i is the $(1-\alpha)\times 100\%$ quantile of the radial density of the device location. Specify this distribution as a symmetric bivariate Gaussian centered at the true device location (μ_x, μ_y) with covariance matrix $\sigma_i^2 I$, where I is the 2×2 identity matrix. Then (X_i, Y_i) has density

$$f(x, y|\sigma_i^2) = \frac{1}{2\pi\sigma_i^2}\exp\left[-\frac{1}{2\sigma_i^2}\left((x-\mu_x)^2+(y-\mu_y)^2\right)\right].$$

If $R_i = r_i$ is the horizontal uncertainty associated with the $(1-\alpha)\times 100\%$ quantile radial density level set of the point i, then $r_i = \sigma_i\Phi^{-1}(1-\alpha)$, where $\Phi^{-1}(\cdot)$ is the standard normal quantile function. We can therefore

estimate the variance σ_i^2 by

$$\hat{\sigma}_i^2 = r_i^2/(\Phi^{-1}(1-\alpha))^2. \tag{1}$$

In this paper, we use $\alpha = 0.05$.

Define the Euclidean distance between the reported location of two points (X_i, Y_i, R_i) and (X_j, Y_j, R_j) as

$$D_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}$$

and fix a distance $\epsilon > 0$. In this paper, we use ϵ equal to two meters. We want to evaluate the probability that points i and j are within ϵ meters of one another. This probability can be expressed as

$$\begin{aligned} \Pr(D_{ij} \leq \epsilon) &= \Pr\left(\sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \leq \epsilon\right) \\ &= \Pr\left((X_i - X_j)^2 + (Y_i - Y_j)^2 \leq \epsilon^2\right) \\ &= \Pr\left(\frac{(X_i - X_j)^2 + (Y_i - Y_j)^2}{\sigma_i^2 + \sigma_j^2} \leq \frac{\epsilon^2}{\sigma_i^2 + \sigma_j^2}\right). \end{aligned} \tag{2}$$

Now under the assumption that (X_i, Y_i) and (X_j, Y_j) have independent bivariate Gaussian distribution, the variance-scaled quantity

$$\frac{(X_i - X_j)^2 + (Y_i - Y_j)^2}{\sigma_i^2 + \sigma_j^2} \tag{3}$$

follows the non-central chi-square distribution with 2 degrees of freedom and non-centrality parameter

$$\frac{(\mu_{xi} - \mu_{xj})^2 + (\mu_{yi} - \mu_{yj})^2}{\sigma_i^2 + \sigma_j^2}. \tag{4}$$

Since the true device locations and variances in (4) are not observed, we substitute the observed device locations X_i, Y_i, X_i, and Y_j, as well as the estimated variances $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$ computed from (1). Because the variance-scaled squared distance (3) follows the non-central Chi-square distribution, the probability that the two devices are within two meters, $D_{ij} \leq 2$, can be computed using standard statistical software.

In reality the Earth is not a plane and the Euclidean distance D_{ij} is shorter than the true distance between i and j on the surface of the Earth. But for distant points or those whose uncertainty radius is large, it is necessary to evaluate longer distances on the surface of the Earth. We therefore substitute the Haversine distance [1] for the Euclidean distance D_{ij} in the calculation above. The resulting Gaussian approximation is useful for small geodesic distances because we are interested in points that are less than two meters apart.

To describe computation of the contact rate, let $Z_i(t) = (X_i(t), Y_i(t), R_i(t))$ be the location and corresponding horizontal uncertainty radius for device i at time t. A *potential contact* between devices i and j at time t occurs when the locations of the two devices $Z_i(t)$ and $Z_j(t)$ are stationary and nearby. Let $D_{ij}(t)$ be the computed distance between the two points i and j. When a potential contact occurs between i and j at time t, let

$$P_{ij}(t) = \Pr(D_{ij}(t) \leq \epsilon)$$

be the probability that these devices are within ϵ meters of each other. Let A_{ad} be the set of pairs of devices for which a potential contact event occurred within area a on day d. For a potential contact between a pair $\{i, j\}$, let t_{ij} be the time of the potential contact. In area a on day d, the expected number of contacts is the sum of the probabilities of contact, across every potential contact event. We compute two contact rates for each area a and day d. First, we aggregate contact probabilities by the area in which the contact occurred. The contact rate by region of contact is

$$C_{ad}^{\text{loc}} = \sum_{\{i,j\} \in A_{ad}} P_{ij}(t_{ij}). \tag{5}$$

Next, we aggregate contacts by the region (town) of the device's primary dwell location. Let A be the set of all regions and let $h(j)$ be the primary dwell region of device j. The device home contact rate is

$$C_{ad}^{\text{home}} = \sum_{b \in A} \sum_{\{i,j\} \in A_{bd}} P_{ij}(t_{ij}) \mathbb{1}\{h(i) = a \text{ or } h(j) = a\}, \tag{6}$$

where the indicator function $\mathbb{1}\{\cdot\}$ is 1 if its argument is true, and 0 otherwise.

1.3 Community case definition

Laboratory tests for SARS-CoV-2 and cases of COVID-19 are reportable to the Connecticut Department of Public Health [2]. A confirmed case of COVID-19 is defined as detection of severe acute respiratory syndrome coronavirus 2 ribonucleic acid (SARS-CoV-2 RNA) in a clinical or autopsy specimen using a molecular amplification test [3]. In this analysis, COVID-19 cases and SARS-CoV-2 tests reported to the Connecticut Department of Public Health among persons residing in long-term care facilities, managed residential communities (e.g., assisted living facilities) or correctional institutions were excluded. Cases and tests among staff working at these locations were not excluded.

The Connecticut Department of Public Health used data from COVID-19 case report forms and laboratory test

reports to identify cases and tests among residents of congregate settings as follows. For COVID-19 cases, public health follow-up is dependent on whether the person resides in a congregate setting (facility does contact tracing) or not (local and state public health do contact tracing). Therefore, the Connecticut COVID-19 case report form includes a field for healthcare providers to indicate whether the patient resided in a congregate setting, and if so, the type of setting. Because case report forms are not received for all cases, this information might also be completed during case investigations by state or local health department staff. For this analysis, cases among congregate setting residents were excluded by using these case report form data.

To identify tests among congregate setting residents, the Connecticut DPH used a geocoding process because residence in a congregate setting is not included in the laboratory result report form, and the number of tests is too large for public health staff to review individual records. This means that the likelihood of misclassification of congregate setting residence among cases is lower than the likelihood of misclassification for tests.

The home address of each person tested is geocoded by the Connecticut Department of Public Health by using a custom composite locator that first attempts to match on the parcel, then street map data. Approximately 90% of test reports are successfully geocoded. To identify tests performed among residents of congregate settings, geocoded results are compared to a spatial layer of known long term care facilities, managed residential communities, and correctional institutions. Records geocoded to within a 150 foot buffer around these locations are categorized as tests performed among residents of congregate settings.

After the steps above were used to exclude cases and tests among congregate setting residents, the Connecticut Department of Public Health tabulated the aggregate count of cases and tests by town for each day, which were used in this analysis. No individually identifiable information was used in the analysis reported here, nor were any linkages made between mobile device data and cases and tests reported to the Connecticut Department of Public Health.

1.4 Data privacy and confidentiality

The mobile device location data contain no individually identifying meta-data. After computing the probability of contact for each potential contact event, we sum contact probabilities by day and by town using (5) and (6). No individual device geolocation data are retained or used in the analysis presented in this paper.

1.5 Transmission model for prediction of COVID-19 infections, cases, hospitalizations, and deaths

We developed a county-stratified deterministic model based on the susceptible-exposed-infective-removed (SEIR) framework to represent COVID-19 transmission and predict case counts in Connecticut using the contact rate. Here we provide a brief description of the model and calibration approach. A detailed description of the model is given in [4]. In the model, the population of each county is divided into 10 compartments. Individuals begin in the susceptible (S) compartment. Exposed individuals (E) may develop either asymptomatic (A), mild (I_M), or severe (I_S) infection. Asymptomatic and mild infections resolve without hospitalization and do not lead to death. Mild symptomatic cases self-isolate (move to R_M) shortly after development of symptoms, and transition to recovery (R) when infectiousness ceases. All severe cases require hospitalization (H) unless hospitalization capacity is exhausted, in which case they transition to $\bar{H}$ representing hospital overflow, then to recovery (R) or death (D). The model assumes a closed population and does not capture non-COVID-19 deaths. Let N_i be the population size of county i and J_i the set of counties adjacent to county i. Let $C^{(i)}$ represent hospitalization capacity in county i, which may vary over time. Transmission dynamics for county i are given by the following system of ordinary differential equations (ODE):

$$
\begin{aligned}
\frac{\mathrm{d}S^{(i)}}{\mathrm{d}t} &= -\beta S^{(i)} \left[(1-k_n) \frac{I_M^{(i)} + I_S^{(i)} + k_A A^{(i)}}{N_i} + \frac{k_n}{|J_i|} \sum_{j \in J_i} \frac{I_M^{(j)} + I_S^{(j)} + k_A A^{(j)}}{N_j} \right] \\
\frac{\mathrm{d}E^{(i)}}{\mathrm{d}t} &= \beta S^{(i)} \left[(1-k_n) \frac{I_M^{(i)} + I_S^{(i)} + k_A A^{(i)}}{N_i} + \frac{k_n}{|J_i|} \sum_{j \in J_i} \frac{I_M^{(j)} + I_S^{(j)} + k_A A^{(j)}}{N_j} \right] - \delta E^{(i)} \\
\frac{\mathrm{d}A^{(i)}}{\mathrm{d}t} &= q_A \delta E^{(i)} - \alpha_A A^{(i)} \\
\frac{\mathrm{d}I_M^{(i)}}{\mathrm{d}t} &= q_{I_M} \delta E^{(i)} - \alpha_{I_M} I_M^{(i)} \\
\frac{\mathrm{d}R_M^{(i)}}{\mathrm{d}t} &= \alpha_{I_M} I_M^{(i)} - \gamma_{R_M} R_M^{(i)} \\
\frac{\mathrm{d}I_S^{(i)}}{\mathrm{d}t} &= q_{I_S} \delta E^{(i)} - \alpha_{I_S} I_S^{(i)} \\
\frac{\mathrm{d}H^{(i)}}{\mathrm{d}t} &= (1-\eta^{(i)}) \alpha_{I_S} I_S^{(i)} - \gamma_H H^{(i)} \\
\frac{\mathrm{d}\bar{H}^{(i)}}{\mathrm{d}t} &= \eta^{(i)} \alpha_{I_S} I_S^{(i)} - \gamma_{\bar{H}} \bar{H}^{(i)} \\
\frac{\mathrm{d}D^{(i)}}{\mathrm{d}t} &= \gamma_H m_H H^{(i)} + \gamma_{\bar{H}} m_{\bar{H}} \bar{H}^{(i)} \\
\frac{\mathrm{d}R^{(i)}}{\mathrm{d}t} &= \alpha_A A^{(i)} + \gamma_{R_M} R_M^{(i)} + \gamma_H (1-m_H) H^{(i)} + \gamma_{\bar{H}} (1-m_{\bar{H}}) \bar{H}^{(i)},
\end{aligned}
\tag{7}
$$

where $q_{I_M} = 1 - q_A - q_{I_S}$. The function $\eta^{(i)} = \left[1 + \exp(0.5(C^{(i)} - H^{(i)}))\right]^{-1}$ is a "soft" hospitalization capacity overflow function. The models captures infection "community" transmission in non-congregate settings, and excludes cases and deaths occurring in settings like nursing homes and prisons. We assume that recovered individuals remain immune to reinfection for the duration of the study period. The analysis was performed using the R statistical computing environment [5]. We used package deSolve to perform numerical integration of the ODE system [6].

1.5.1 Time-varying model parameters

Recognizing that many of the model parameters are unlikely to be constant over time, we represent the most critical parameters as functions of time. These parameters include: transmission parameter β, rates α_{I_M} and α_A (reciprocals of average duration of infectiousness among mildly symptomatic and asymptomatic individuals), q_{I_S} - proportion of infections, which are severe and require hospitalization, rate of hospital discharge γ_H, and hospital case fatality ratio m_H.

We use data on close contact to parameterize temporal dynamics of transmission parameter β:

$$\beta(t) = \beta_0 M_{\text{contact}}(t) \exp[B(t)],$$

where $M_{\text{contact}}(t)$ is a measure of close contact at time t relative to the pre-epidemic level, and $\exp[B(t)]$ is a function that approximates residual changes in transmission parameter β that are not explained by changes in close contact or other time-varying parameters. $B(t)$ is a smooth function obtained by applying spline smoothing on a piecewise linear function $B^*(t)$, where $B^*(t)$ is modeled with $B^*(w) = \epsilon_{[(w-t_0)/14]}$ defined on bi-weekly knots $w = \{t_0, t_0 + 14, t_0 + 28, \ldots\}$ over the observation period and linearly imputed between the knots. We model the vector of random effects $\boldsymbol{\epsilon}$ using a random walk of order one:

$$\epsilon_0 = 0, \quad \epsilon_i | \epsilon_{i-1} \sim \mathcal{N}(\epsilon_{i-1}, \sigma_\epsilon^2).$$

For the hyperparameter σ_ϵ we let

$$\frac{1}{\sigma_\epsilon^2} \sim \text{Gamma}(a_\epsilon, b_\epsilon),$$

with shape parameter $a_\epsilon = 2.5$ and rate parameter $b_\epsilon = 0.1$.

Rates of isolation or recovery α_{I_M} and α_A are assumed to depend on the testing volume. Widespread testing is assumed to identify infectious cases sooner, resulting in faster isolation and shorter duration of infectiousness. Via its relationship to age distribution of cases, probability of severe infection is assumed to change over time, likely due to higher adoption of social distancing measures by individuals with worse prognosis [7]. Time-varying rate of hospital discharge γ_H was obtained from the Connecticut Hospital Association (CHA) based on the analysis of insurance claims data. Temporal variation in the hospital case fatality ratio m_H was estimated based on data on daily hospital deaths and admissions.

Table 1 summarizes the parameters in the model and their meaning. We fix some of the model parameters whose values can be estimated based on prior research.

1.5.2 Calibration data

We calibrate the joint distribution of model parameters at the state level using the observed dynamics of confirmed COVID-19 hospitalizations census, cumulative COVID-19 hospitalizations, and cumulative number of deaths among hospitalized cases, which was obtained from the Connecticut Hospital Association [35]. In addition to close contact data, we use the following data to inform time-varying model parameters: average hospital length of stay by month (source: CHA [35]), testing volume data, and data on the age distribution of confirmed cases (source: Connecticut Department of Public Health (CT DPH) daily reports [36]). We calculated non-institutionalized county-level population and age structure in Connecticut using the 2014–2018 estimates of the American Community Survey [37]. Daily total available hospital beds (including occupied) in each county were obtained from the Connecticut Hospital Association/CHIMEData [35, 38] and used as hospitalization capacity values on a given date.

Since available hospitalizations data does not disaggregate by the patient's place of residence at the time of diagnosis or hospitalization, we estimate the time-varying distribution of hospitalizations and hospital deaths coming from congregate and non-congregate settings based on the daily distribution of COVID-19 death counts occurring in hospitals by the type of residence at the time of diagnosis (congregate vs. non-congregate) and a case fatality ratio among hospitalized residents of congregate settings. These data were obtained from Connecticut Department of Public Health.

Parameter	Meaning	Value	Source
q_A	Proportion of infections that are asymptomatic	Calibrated, prior mean = 0.4	[8–10]
$q_{I_S,0}/(1-q_A)$	Proportion of infections that are severe among symptomatic	0.065	[11–13]
β_0	Transmission parameter at baseline	Calibrated, prior mean = 1	Diffuse prior assumed
δ	1/duration of latency	1/4	Of 6 days incubation, 2 days are presymptomatic [10, 12, 14–23]
$\alpha_{A,0}$	1/duration of asymptomatic infectiousness at baseline	1/7	[16, 24–29]
k_A	Relative infectiousness of asymptomatic compared to symptomatic	0.5	[9, 15, 26, 30, 31]
$\alpha_{I_M,0}$	1/duration of symptomatic infectiousness at baseline	1/4	[11, 12, 14]; of 4 days, 2 days are presymptomatic [15–17]
γ_{R_M}	1/duration of isolation among mild symptomatic	1/7	[24, 28, 29, 32]
α_{I_S}	1/time between infectiousness onset and hospitalization	Calibrated, prior mean = 1/9	[15–17, 33, 34]
γ_H	1/length of hospital stay	Time-varying	CHA data on average hospital length of stay by month [35]
$m_{H,0}$	Hospital case fatality ratio at baseline	Calibrated, prior mean = 0.13	Estimated from CHA COVID-19 hospitalization data [35]
k_n	Proportion of contact between adjacent counties	0.015	Assumed
E_0	Number of exposed individuals at the epidemic onset	Calibrated, prior mean = 150	Assumed
L_H	Hospitalization reporting lag	Calibrated, uniform prior on integers between 5 - 14 days	Assumed
L_D	Death reporting lag	Calibrated, uniform prior on integers between 5 - 14 days	Assumed

Table 1: Transmission model parameters. Fixed parameters are drawn from the cited sources. Calibrated parameters are estimated using the Bayesian procedure described below. Durations are measured in days.

1.5.3 Model calibration and Bayesian posterior inference

We calibrate a joint posterior distribution of model parameters $\boldsymbol{\theta} = (\beta_0, q_A, \alpha_{I_S}, m_{H,0}, E_0, L_H, L_D, \boldsymbol{\epsilon})$ and hyperparameters $\boldsymbol{\sigma} = (\sigma_h, \sigma_u, \sigma_d, \sigma_\epsilon)$ to the observed data using a Bayesian approach with a Gaussian likelihood. We also accommodate reporting lags in observed hospitalizations (L_H) and hospital deaths (L_D). The distributions of hospitalizations census, cumulative hospitalizations, and hospital deaths are given by:

$$h(t) \sim \mathcal{N}\left(H(t, L_H, \boldsymbol{\theta}), \sigma_h^2\right), \tag{8}$$

$$u(t) \sim \mathcal{N}\left(U(t, L_H, \boldsymbol{\theta}), \sigma_u^2\right), \tag{9}$$

$$d(t) \sim \mathcal{N}\left(D(t, L_D, \boldsymbol{\theta}), \sigma_d^2\right), \tag{10}$$

where $H(t, L_H, \boldsymbol{\theta})$, $U(t, L_H, \boldsymbol{\theta})$, and $D(t, L_D, \boldsymbol{\theta})$ are model-projected hospitalizations census (lagged by L_H), cumulative hospitalizations (lagged by L_H), and cumulative hospital deaths (lagged by L_D) at time t under parameter values $\boldsymbol{\theta}$. Reporting lags L_H and L_D are correlated with other model parameters, including latency period, time between infection and hospitalization, time between infection and death, and length of hospital stay, therefore L_H and L_D should not be strictly interpreted as reporting lags. We assume the date of epidemic onset to be February 16, 2020 - 21 days before the first case was officially confirmed in Connecticut on March 8th, 2020, and initialize the model with E_0 exposed individuals at the time of epidemic onset and set the initial size of all downstream compartments to be zero. We assume a prior distribution on E_0 and calibrate it along with other model parameters. The county-level distribution of E_0 is fixed and was estimated based on the county population size and dates of first registered case and death in each county. We put the same independent Inv-Gamma(a, b) prior on the three hyperparameters σ_h^2, σ_u^2, and σ_d^2 with $a = 50$ and $b = 5 \times 10^7$.

We construct the joint posterior distribution over unknown parameters $(\boldsymbol{\theta}, \boldsymbol{\sigma})$ as

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\sigma} | h(t), u(t), d(t)) \propto p(\boldsymbol{\theta}) p(\boldsymbol{\sigma}) & \prod_{t \in t_H} \left[p\big(h(t) | H(t, L_H, \boldsymbol{\theta}), \sigma_h\big)\right]^{w_h z(t)} \\ & \prod_{t \in t_U} \left[p\big(u(t) | U(t, L_H, \boldsymbol{\theta}), \sigma_u\big)\right]^{w_u z(t)} \\ & \prod_{t \in t_D} \left[p\big(d(t) | D(t, L_D, \boldsymbol{\theta}), \sigma_d\big)\right]^{w_d z(t)}. \end{aligned} \tag{11}$$

Each likelihood term is weighted by $z(t)$ (observation weight at time t) times the weight assigned to a respective time series. We set $w_h = 0.89$, $w_u = 0.01$, and $w_d = 0.1$. We place a large weight on the hospitalizations

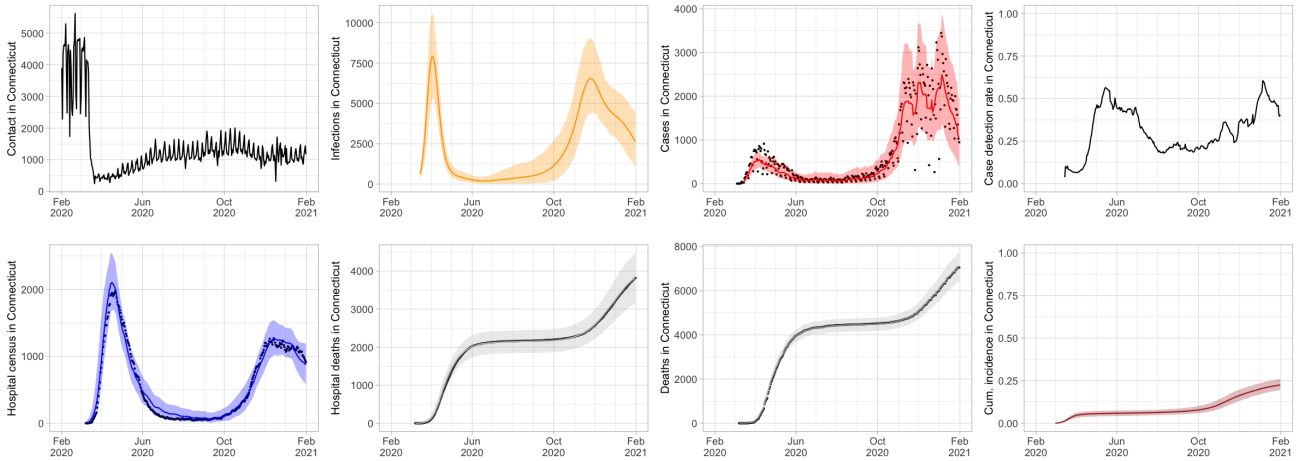

Figure 1: Connecticut statewide data and calibrated model projections. Top row from left to right: statewide contact, projected daily infections, projected (red line) and observed (black dots) cases, estimated statewide case detection rate. Bottom row from left to right: projected and observed hospital census, projected and observed hospital deaths, projected and observed total deaths, estimated cumulative incidence. Model projections are shown as lines with 95% posterior intervals. Observed data is shown as black dots.

census, since this time series is most sensitive to changes in epidemic dynamics, and a small weight on cumulative hospitalizations, since it measures a feature that is related to hospitalizations census. The range of observation times differ for different time series. For hospitalizations census and deaths, observation times start with the first non-zero observation. For cumulative hospitalizations, observation times start on May 29, 2020 when this indicator started being reported routinely. The last observation in all three time series is from February 15, 2021.

Sampling from the joint posterior distribution of (θ, σ) given in (11) is performed using Markov Chain Monte Carlo (MCMC). We employ a hybrid algorithm that combines elliptical slice sampling (ESS) [39], Gibbs sampling, and Metropolis-Hastings sampling with random walk proposals. Details of the algorithm implementation are provided in [4]. Bottom row of Figure 1 shows statewide model fit to hospital census, hospital deaths, and total deaths.

1.5.4 Model projections for case counts at the town level

To produce town-level projections, we use parameters estimated from the state-level joint posterior distribution and town-level inputs. Town-level data consist of the normalized close contact function, town population size, and town-specific initial model compartment conditions. Normalized close contact at the town level is obtained by adding an intercept to the town level contact that equals the median difference between the state and the

town-level contact for all dates t after April 7, 2020. This step enables the application of a random effects vector ϵ estimated at the state-level to the town-level contact data. The normalization approach preserves town-level contact dynamics and contact levels in each town early in the epidemic and immediately after the implementation of social distancing measures. Town-specific initial conditions were computed from the state-level posterior distribution of E_0 based on the town population size and its proximity to New York City.

We simulate epidemic dynamics for each of the 169 Connecticut towns using the model given in (7), modified to include only data from a single town. Using the model-projected dynamics of lagged daily infections within the town, we estimate the time-varying COVID-19 case-finding rate for each town using a log-linear regression model:

$$\mathbb{E}[\log(\text{Cases}_{it})] = \alpha_0 + \alpha_i + \gamma_t + \beta \log(\text{Infections}_{i(t-l)}),$$

where Cases_{it} is the 7-day moving average of reported COVID-19 cases in town i at time t, α_i is a town effect, γ_t is a time effect measured daily and $\text{Infections}_{i(t-l)}$ is the model-projected number of daily infections in town i at time $t - l$, and we set $l = 14$ days. As a final step to produce model-based estimates of daily cases in each of 169 Connecticut towns, we truncate regression-estimated town-level case-finding rate at 1 and apply this estimated rate to model-projected daily infections. The top row of Figure 1 shows statewide contact measure, projected infections, estimated and observed non-congregate COVID-19 case counts, and estimated case-detection rate. Estimates of statewide case counts is obtained by summing up estimated cases across 169 Connecticut towns.

1.6 Space-time regression model

To assess the relationship between close interpersonal contact and COVID-19 cases without assumptions about the dynamics of transmission, we develop a statistical model for describing the associations between town-level contact in previous days and current day COVID-19 cases using a distributed lag negative binomial regression framework. We emphasize that this "phenomenological" regression model does not use any SEIR-type assumptions about the relationship of contact, infections, and cases. Our purpose is to validate the usefulness of contact data in prediction of future case counts throughout the state.

In the model, each town has a unique set of parameters that describe the lagged associations with past contact, as we hypothesize that there may be spatial differences in these associations due to town-level features (e.g., population density). Spatial correlation between these parameters is also modeled since towns close to each

other may exhibit similar spatial patterns. Specifically, the model is given as

$$\begin{aligned} Y_{it}|r, p_{it} &\overset{\text{ind}}{\sim} \text{Negative Binomial}\left(r, p_{it}\right),\ i = 1, \ldots, n,\ t = 1, \ldots, m, \\ \text{logit}\left(p_{it}\right) &= \mathbf{x}_{it}^{\mathsf{T}}\boldsymbol{\beta} + \sum_{j=1}^{d} \mathsf{z}_{i,t-j}\theta_{ij} + \eta_i + \lambda_t \end{aligned} \tag{12}$$

where Y_{it} is the number of COVID-19 cases recorded in town i on day t, n is the total number of towns in the analysis ($n = 169$), and m is the total number of included days of data ($m = 351$). In our application of the model, the first day of analysis ($t = 1$) occurs on February 29th, 2020 and the final day ($t = 351$) occurs on February 13th, 2021.

In (12), we use the logit link function to connect the underlying probability that controls the expected number of daily cases in a location to the covariates of interest, where $\mathbf{x}_{it}$ is a vector of town-specific non-time varying covariates (i.e., median household income, total population, percent of the population living below the poverty level, percent of the population that is non-Hispanic White, percent of the population with some college education, and percent of the population that is $\geq$ 65 years) along with a day-of-the-week indicator to account for systematic case reporting patterns across time; $\boldsymbol{\beta}$ is the corresponding vector of regression parameters. Contact in town i on day t is denoted as z_{it}, and in our application of the model we include contact up to $d = 28$ days (i.e., 4 weeks) in the past. Town-specific (η_i) and time-specific (λ_t) random effects are included to account for separable spatial and temporal correlation, respectively, in the case data.

The θ_{ij} parameters are town- and lag-specific, and describe the association between contact j days earlier in the town and current day cases. The collection of all lag parameters from town i, $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{i28})^{\mathsf{T}}$, are decomposed into "global" and "local" effects to account for similarities in associations shared across all towns and town-specific deviations from that trend such that

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \widetilde{\boldsymbol{\theta}}_i,\ i = 1, \ldots, n$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{28})^{\mathsf{T}}$ represents the vector of parameters shared across all towns and $\widetilde{\boldsymbol{\theta}}_i = \left(\widetilde{\theta}_{i1}, \ldots, \widetilde{\theta}_{i28}\right)^{\mathsf{T}}$ are town-specific deviations from the global pattern. We anticipate that the global distributed lag parameters closer together in lag time will have similar behavior and model this directly by specifying

$$\boldsymbol{\theta}|\tau_\theta^2, \phi \sim \text{MVN}\left(\mathbf{0}_{28},\ \tau_\theta^2 \Sigma\left(\phi\right)\right)$$

where MVN represents the multivariate normal distribution, $\mathbf{0}_{28}$ is a vector of 28 zeros, and the correlation between the global lag parameters is defined by

$$\text{Corr}\left(\theta_j, \theta_{j'}\right) = \Sigma\left(\phi\right)_{jj'} = \exp\left\{-\phi|j - j'|\right\}.$$

The variance and smoothness of the process are described by τ_θ^2 and ϕ, respectively, where smaller values of ϕ suggest higher correlation between parameters across the lags.

Spatial and cross-correlations between the town-specific lag parameters ($\widetilde{\boldsymbol{\theta}}_i$) and town-specific intercepts (η_i) are modeled using a multivariate conditional autoregressive model [40] given as

$$\boldsymbol{\psi}_i = \left(\eta_i, \widetilde{\boldsymbol{\theta}}_i^{\mathsf{T}}\right)^{\mathsf{T}} | \boldsymbol{\psi}_{-i}, \rho, \Omega \overset{\text{ind}}{\sim} \text{MVN}\left(\frac{\rho \sum_{j=1}^n w_{ij}\boldsymbol{\psi}_j}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho}, \frac{\Omega}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho}\right), \; i = 1, \ldots, n$$

where $\boldsymbol{\psi}_{-i}$ is the entire set of $\boldsymbol{\psi}_j$ vectors with $\boldsymbol{\psi}_i$ removed and w_{ij} is an indicator describing whether towns i and j share a common border or not. As is standard practice, towns are not considered to be neighbors of themselves, resulting in $w_{ii} = 0$ for all i. This model specifies that *a priori*, the set of parameters from a specific town have a multivariate normal distribution with a mean vector equal to a weighted mean of its neighbors vectors. The amount of spatial similarity between towns is controlled by $\rho \in (0, 1)$ where a value of ρ close to zero indicates near independence and close to one results in strong spatial smoothing. The cross-correlation between parameters within a given town is described by the unstructured covariance matrix Ω.

Finally, λ_t represents a flexible time trend that is common to all towns in the analysis and is defined using an autoregressive model such that

$$\lambda_t | \lambda_{t-1}, \alpha, \tau_\lambda^2 \overset{\text{ind}}{\sim} \text{N}\left(\alpha\lambda_{t-1}, \tau_\lambda^2\right), \; t = 1, \ldots, m$$

where $\lambda_0 \equiv 0$ and $\alpha \in (0, 1)$ controls the level of similarity across time. This effect serves to account for temporal correlation between the case counts and describe global patterns of risk that vary across time.

1.6.1 Prior distributions

We complete the model by specifying prior distributions for all introduced parameters. When possible, we opt for weakly informative priors so that the data are the main drivers of the inference. The priors are given

as $r \sim$ Discrete Uniform $[1, 20]$, $\beta_j \overset{\text{iid}}{\sim} \text{N}\left(0, 100^2\right)$ for all j, $\tau_\lambda^2, \tau_\theta^2 \overset{\text{iid}}{\sim}$ Inverse Gamma $(0.01, 0.01)$, $\alpha, \rho \overset{\text{iid}}{\sim}$ Uniform $(0.00, 1.00)$, $\phi \overset{\text{iid}}{\sim}$ Gamma $(1.00, 1.00)$, and $\Omega^{-1} \sim$ Wishart $(30, I_{29})$ where I_{29} is the 29 by 29 identity matrix.

1.6.2 Competing methods

To determine the importance of previous days contact on explaining trends in current day COVID-19 case counts, we also test two competing modeling options. First, we consider the "Global Contact" model, where (12) is modified such that $\boldsymbol{\theta}_i$ is replaced with $\boldsymbol{\theta}$ ($\boldsymbol{\theta}$ previously described). As a result, in Global Contact there are no town-specific lag parameters, only a single set that is shared across all towns. We expect this model to perform poorly compared to the full spatio-temporal model in (12) if there is spatial variability between these associations. Next, we consider the "No Contact" model, which modifies (12) by removing all of the contact information entirely. This model is likely to struggle to explain variability in the case data if prior days contact are important predictors. We note that both of these competing models technically represent subsets of the full spatio-temporal model in (12). For example, when $\widetilde{\boldsymbol{\theta}}_i \equiv \mathbf{0}$, the full model becomes Global Contact and when $\boldsymbol{\theta}_i \equiv \mathbf{0}$ it becomes No Contact. Due to this flexibility, we anticipate that the full model will perform well regardless of the spatio-temporal trends in the data.

We fit each method to the data and compare the findings using multiple methods. First, we use Watanabe–Akaike information criteria (WAIC) [41] to compare the explanatory ability of each method. WAIC is a Bayesian model selection tool that identifies the model, among a small subset of competitors, that has the best balance of fit to the data and complexity, with smaller WAIC values being preferred. Next, we consider a posterior predictive model comparison tool focused on the predictive ability of each method [42]. Specifically, we calculate the posterior mean of the deviance of the negative binomial regression model evaluated at the observed data and data generated from the posterior predictive distribution (from the same design points as the observed data). Smaller values of this metric suggest improved predictive performance of the corresponding method.

1.6.3 Model fitting

We fit all models using Markov chain Monte Carlo sampling techniques, including a Pólya-Gamma latent variable approach for obtaining semi-conjugacy for many of the parameters within the negative binomial regression

model [43]. From each model, we collect 100,000 samples from the joint posterior distribution after removing the first 10,000 iterations prior to convergence of the model. We further thin the samples by a factor of 10 to reduce posterior autocorrelation, resulting in 10,000 samples with which to make posterior inference. Convergence of the models was assessed by visually inspecting individual parameter trace plots and calculating Geweke's diagnostic [44] for all relevant parameters. Neither tool suggested any obvious signs of non-convergence.

2 Supplementary Text

2.1 Mobile device coverage and representativeness in Connecticut

We investigated the coverage of mobile devices in the sample by town population, percent who do not identify as "only white", percent without a high school degree, and percent below the poverty level as defined by the federal government's official poverty threshold, which accounts for household size, number of children, and age of householder. Demographic data were obtained from the American Community Survey [37, 45]. Figure 2 shows that device coverage tracks population size throughout the state, but coverage decreases slightly with the percent non-white, the percent that lack a high school degree, and percent below the poverty level. The town of Union is the smallest town in Connecticut, with a total population of 873 people. We do not know whether lower coverage is due to a lower percentage of town residents owning a mobile device, or because of systematic under-coverage of devices from these towns in our sample.

We received mobile device geolocation data in two overlapping batches, using different methods of anonymizing device IDs. In the first batch, covering February 1 – May 31, 2020, we observed a total of 573,452 unique device IDs in Connecticut. The average number of unique device IDs per day was 141,617. Each week an average of 80.5% of the device IDs carried over to the next week. In the second batch, covering May 1, 2020 to January 31, 2021, we observed a total of 788,842 unique device IDs in Connecticut. The average number of unique device IDs per day was 114,946. Each week an average of 77.6% of the device IDs carried over to the next week. Connecticut has a population of 3.565 million people [37, 45]. Contacts occurring within a given town are not captured if we cannot calculate a primary device dwell location for both devices involved in the contact. Individual device IDs might drop out of the sample for several reasons:

- the device leaves the state of Connecticut,

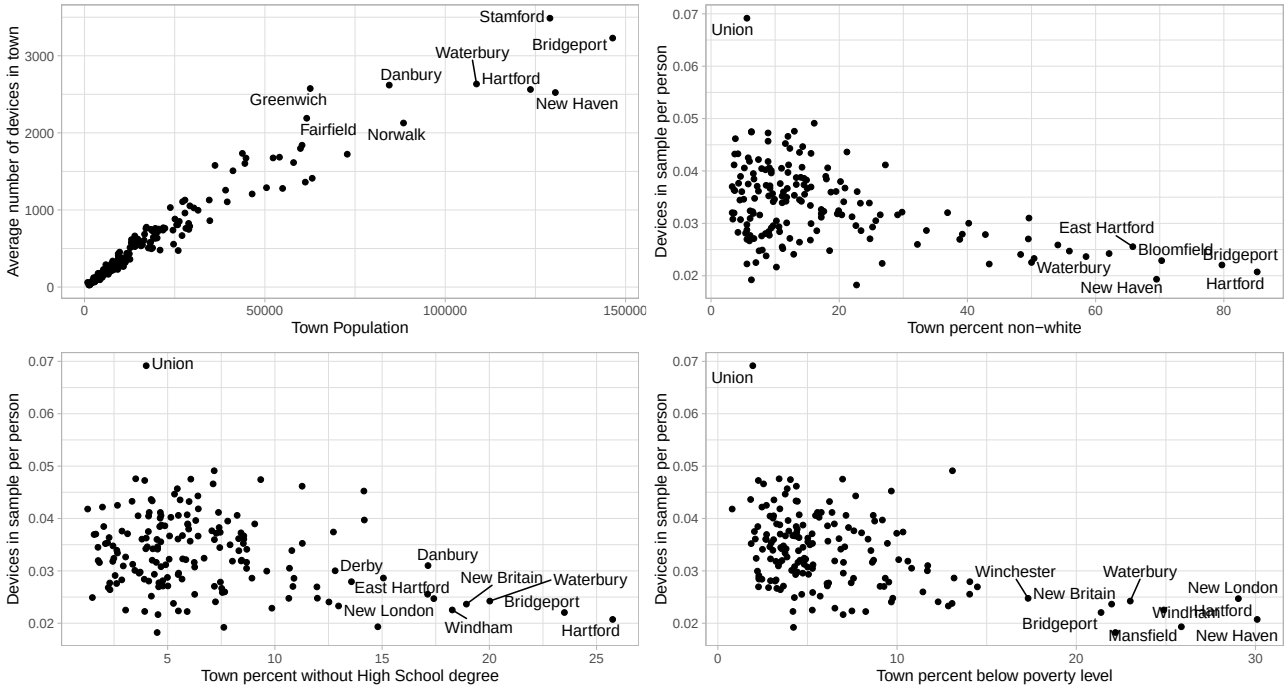

Figure 2: Number of mobile devices in the sample by Connecticut town. Top left: average number of devices in each town by total town population; top right: devices in the sample per person by town percent not identifying as "only white"; bottom left: devices in the sample per person by percent without a high school degree; and bottom right: devices in the sample per person by percent below the poverty level.

- the user uninstalls the applications reporting data to X-Mode,
- the user stops using an application that reports location data only while in use,
- the user rotates their device ID, or
- the user gets a new device with a new advertising ID.

Figure 3 shows the number of unique devices per day in the Connecticut sample.

2.2 Interactive web application

We created an interactive web application to allow users to explore contact patterns in Connecticut over time, available at `https://forrestcrawford.shinyapps.io/ct_social_distancing/`. Figure 4 shows a screen shot of the interactive application. The application uses the Shiny framework [46] for the R computing environment [5], Leaflet for mapping [47], and tigris for shapefiles [48]. The map explorer is based in part on Edward Parker's COVID-19 tracker [49] and code [50].

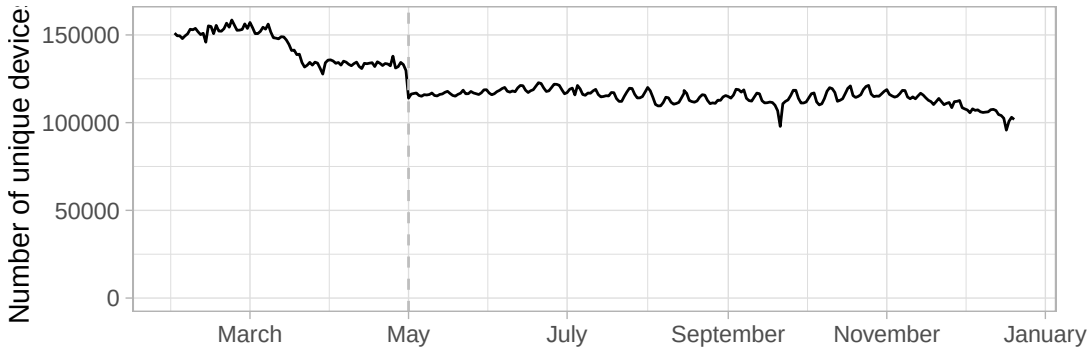

Figure 3: Number of unique mobile device IDs per day in the Connecticut sample. The vertical dashed line on May 1, 2020 corresponds to a change in the unique device ID coding from the data supplier.

The application shows contact by location of contact (5) and by device primary dwell town 6 for each day, at the town and census block group levels. Users can view the contact maps over time from February 1, 2020 to the present, as well as time trends of contact at the state and local levels. The application shows the top contact towns and census block groups throughout Connecticut, as well as points of interest – businesses, schools, hospitals – that help identify block groups.

2.3 Comparison of contact rate to mobility metrics

In order to compare the contact rate described in this paper to other mobility metrics, we acquired Connecticut mobility data from Google [51], Apple [52], Facebook [53], Descartes Labs [54, 55], and Cuebiq [56]. Google and Apple provide public data access, while Facebook and Cuebiq provide data through their respective Data For Good programs. We normalized all metrics to a day-of-week baseline using data from January or February depending on availability, and plot their percent change from baseline from February 2020 through January 2021.

Apple state-level data measures Apple Maps routing requests, categorized as transit, walking, or driving. Map routing requests are a proxy for mobility but might not represent actual trips. Movements for which Apple Maps directions are not needed, such as everyday trips for work, school, or shopping, might not be represented in routing request metrics. Figure 5 shows mobility metrics published by Apple using the day-of-week median during February 2 – February 29, 2020 as a baseline. While transit use remained below baseline during March 2020 through January 2021, driving and walking returned to baseline in June 2020. Driving and walking remained above baseline until November 2020, at which point they returned to near baseline through January 2021.

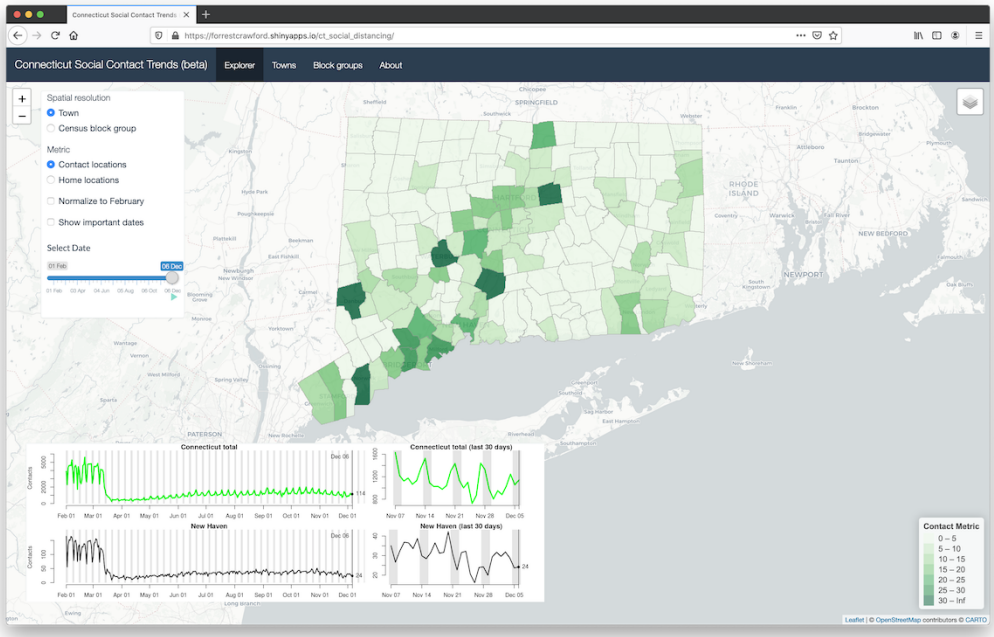

Figure 4: Interactive web application showing contact in Connecticut towns on December 6, 2020. The application can display the locations where contact is occurring or contact by the town of device primary dwell town.

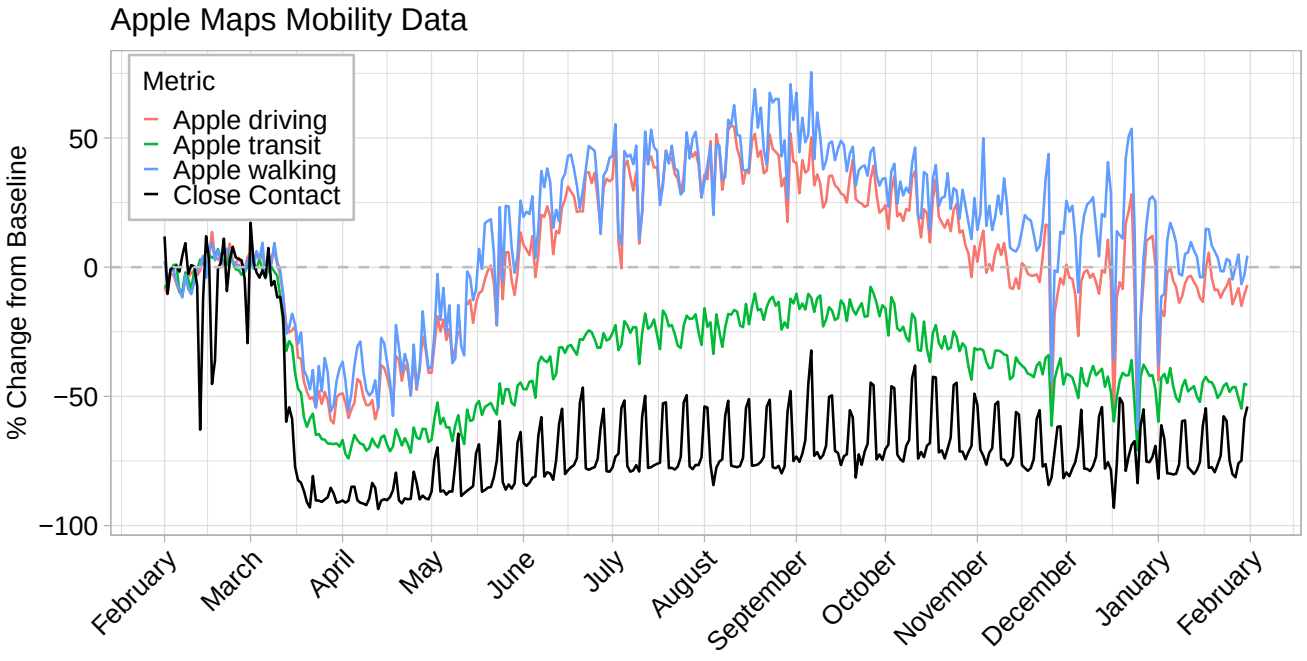

Figure 5: Comparison of Apple Maps mobility metrics to the contactr rate described in this paper during February 1 – January 31, 2021.

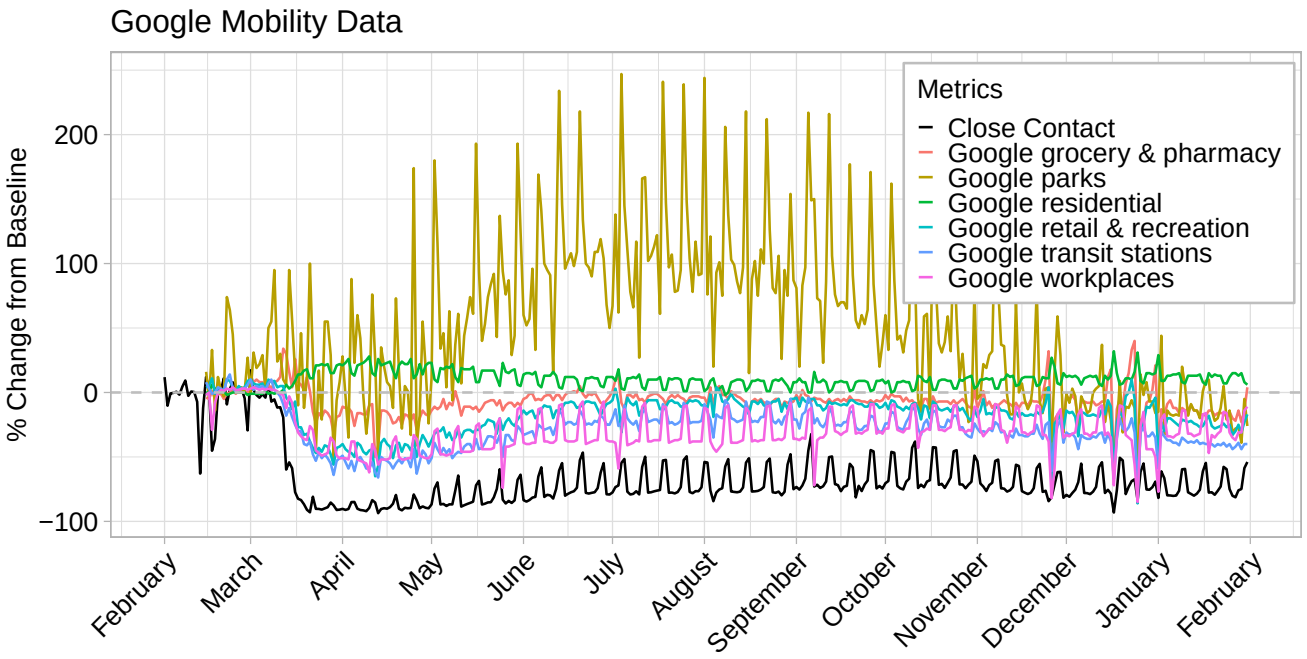

Figure 6: Comparison of Google mobility metrics to the contact rate described in this paper during February 1 – January 31, 2021.

Google state-level mobility data measured visits to areas of interest, categorized as grocery and pharmacy, parks, residential, retail and recreation, transit stations, and workplaces. More detailed information about the definitions of these areas of interest, and the completeness of these categories, is not available. Figure 6 shows mobility metrics published by Google using the day-of-week median from January 3, 2020 to February 6, 2020 as the baseline. All categories other than transit stations and workplaces returned to near baseline levels by summer 2020, and all categories other than residential remained near or below baseline throughout winter 2020.

Facebook county-level mobility data measured the number of 600m by 600m geographic units visited by a device in a day. This metric summarizes how mobile people from different counties are, but might not represent the distance of travel, time away from home, or potential close contacts with others. Figure 7 shows mobility metrics published by Facebook with day-of-week mean during February 2 – February 29, 2020 (excluding February 17) as the baseline. Facebook mobility levels returned to near baseline in all Connecticut counties by July 2020, with little difference between counties. From fall 2020 through January 2021, Facebook mobility levels for all counties decreased to slightly below baseline levels.

Cuebiq county-level mobility data measures a 7 day rolling average of the median distance traveled in a day,

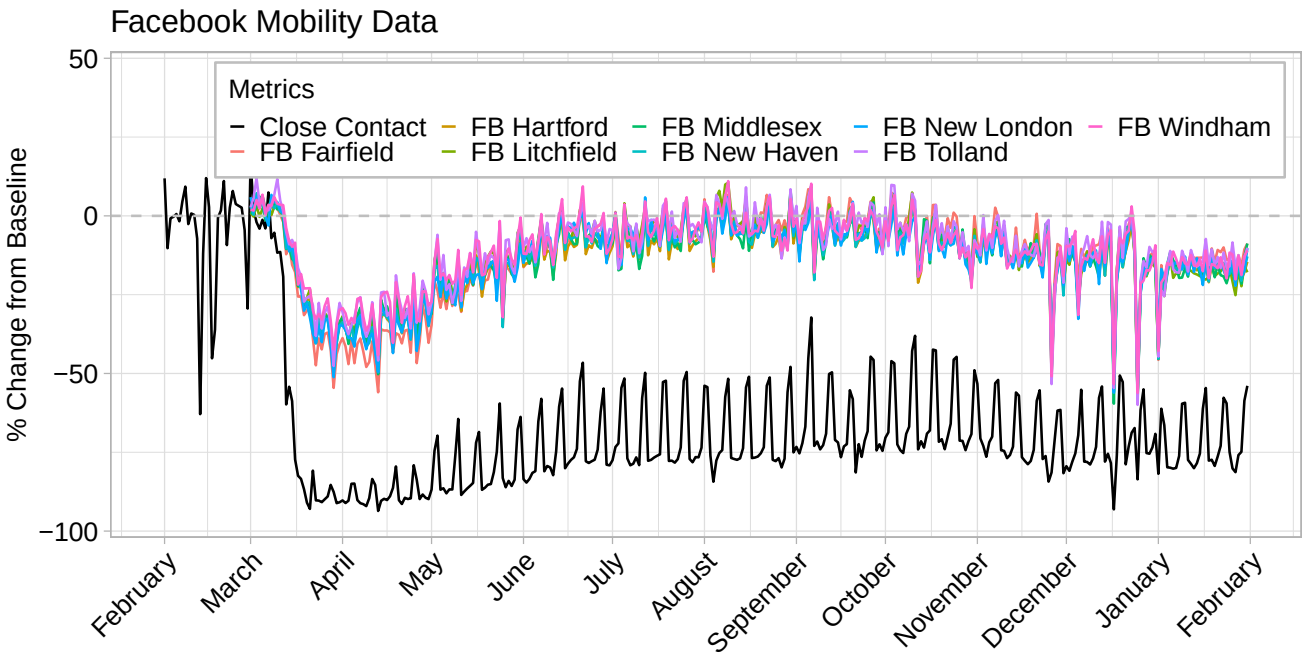

Figure 7: Comparison of Facebook (FB) mobility metrics to the contact rate described in this paper during February 1 – January 31, 2021.

and was available through November 1, 2020 [1]. Figure 8 shows mobility data provided by Cuebiq with day-of-week median during February 2 – February 29, 2020 as the baseline. By July 2020, Cuebiq mobility levels returned to near baseline. Figure 9 shows Cuebiq's metric for "contact", when two or more devices are within 50 feet of each other within five minutes. Information about whether this metric takes spatial error (horizontal uncertainty) into account is not available. In July 2020, Cuebiq contact levels remained further below baseline than the Cuebiq mobility metric. The Cuebiq 50-foot contact metric was closer to baseline than our calculated contact rate during summer and fall 2020.

Finally, Descartes Labs state-level mobility data represents maximum distance devices have moved from the first reported location in a given day. Figure 10 shows the mobility metric provided by Descartes Labs with day-of-week median during February 17 – March 7, 2020 as the baseline. It was exceptional amongst the data sources in that mobility remained notably below baseline during March 2020 – January 2021. However, the percent decline in close contact was consistently larger than the observed percent decline in the Descartes Labs mobility metric.

Every mobility metric we studied returns to baseline faster than the contact rate, though Descartes Labs' mo-

[1] Aggregated mobility data were provided by Cuebiq, a location intelligence and measurement platform. Through its Data for Good program, Cuebiq provides access to aggregated mobility data for academic research and humanitarian initiatives. This first-party data is collected from anonymized users who have opted-in to provide access to their location data anonymously, through a GDPR-compliant framework. It is then aggregated to the census-block group level to provide insights on changes in human mobility over time.

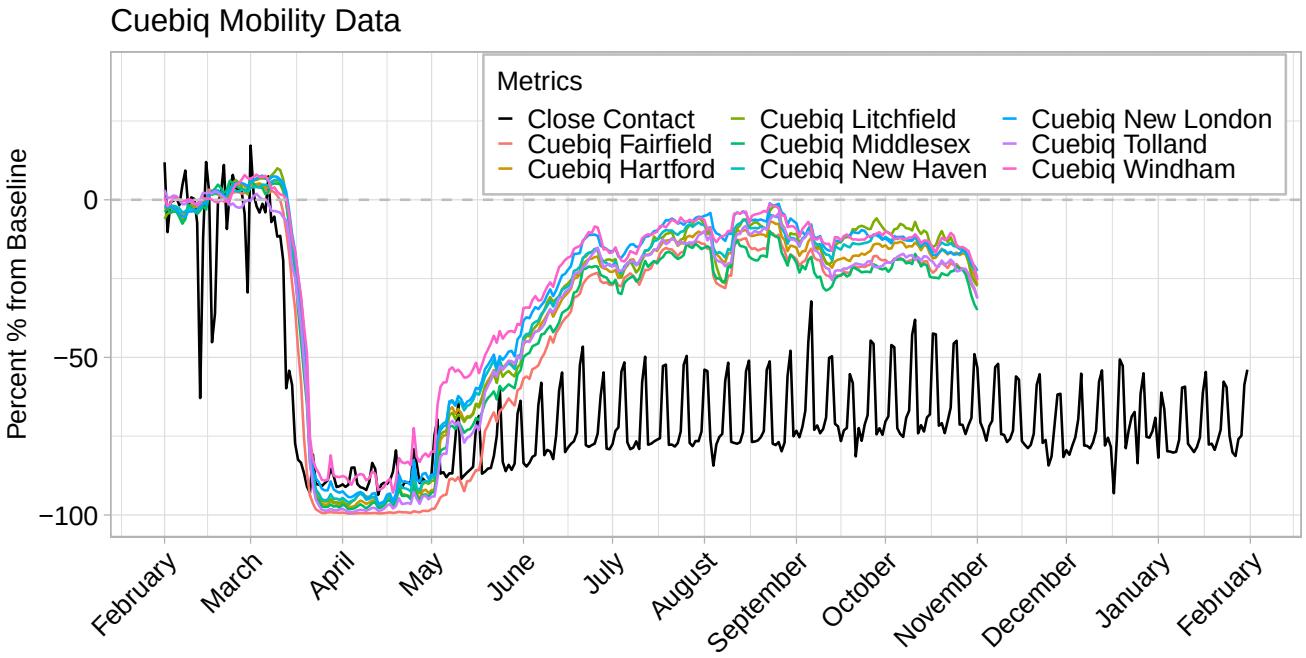

Figure 8: Comparison of Cuebiq mobility metrics to the contact rate described in this paper during February 1 – January 31, 2021. Cuebiq data available through November 1, 2020.

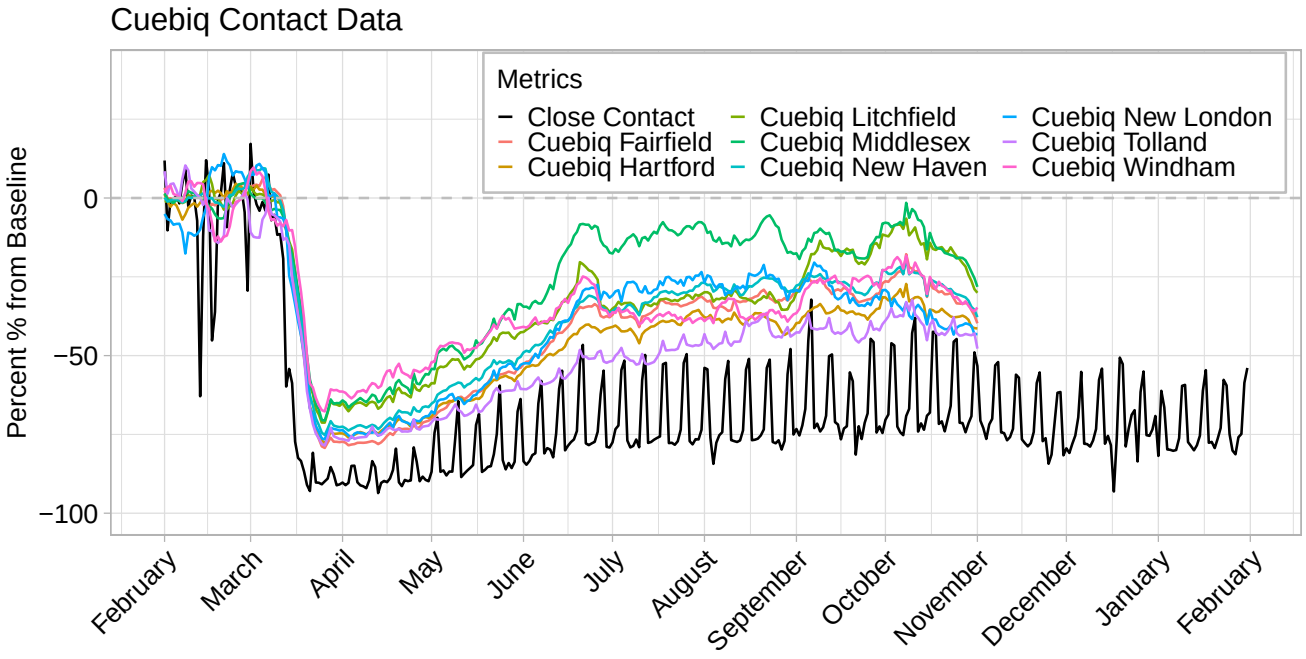

Figure 9: Comparison of the 50-foot Cuebiq contact metric to the contact rate described in this paper during February 1 – January 31, 2021. Cuebiq data available through November 1, 2020.

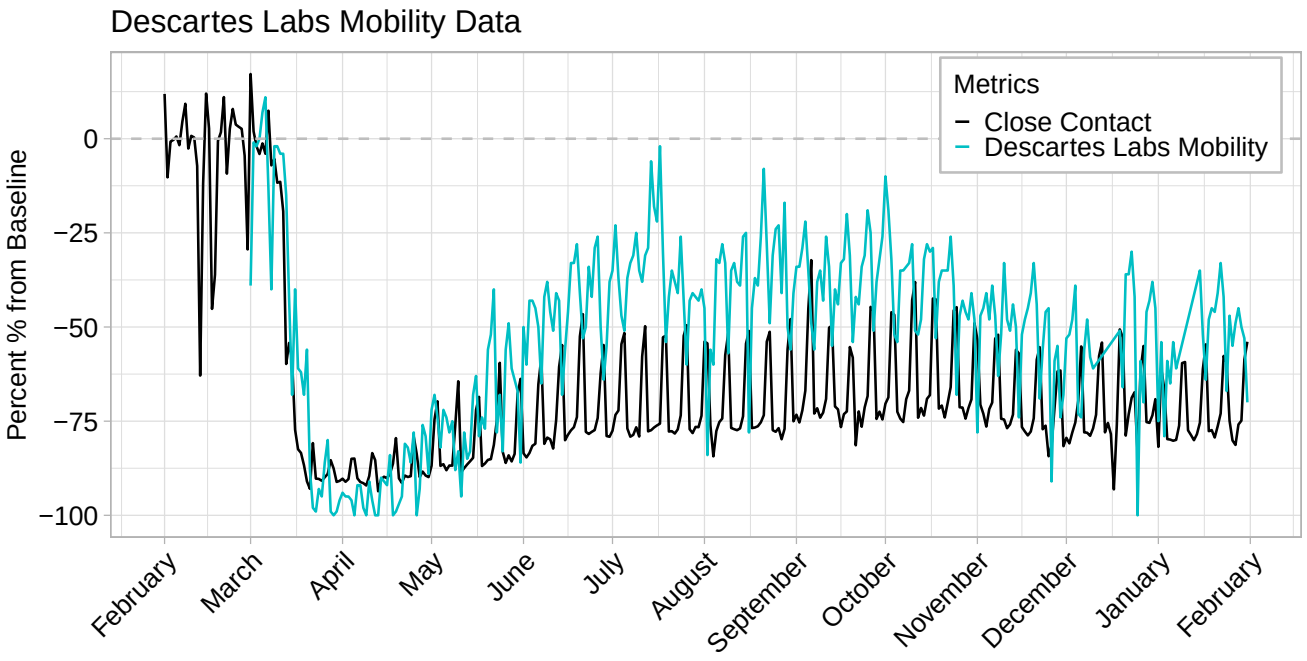

Figure 10: Comparison of Descartes Labs mobility metrics to the contact rate described in this paper during February 1 – January 31, 2021.

bility data provide the closest match to contact when presented on the percent change from February scale. In general, mobility metrics rebounded to February 2020 levels during summer 2020, and remained near this baseline or declined slightly through fall and winter 2020. Close contact, however, remained well below baseline levels through January 2021, suggesting that trends in mobility do not necessarily align with trends in close contact.

2.4 Space-time regression model results

In Table 2, we display the WAIC and posterior predictive deviance model comparison results. WAIC clearly favors the full spatio-temporal method even though the model is more complex than the two competitors as indicated by the effective number of parameters (p_{WAIC}). This improvement in model fit however, overpowers this increased penalization term, leading to an improved overall balance. In terms of prediction, the full model is again preferred and seems to predict data with similar features to the observed data better than the competing models. Global Contact fits and predicts better than No Contact, suggesting that including contact information in some form within the model leads to improvements in both areas. Overall, these results show the importance of including prior day contact information in a flexible manner in order to explain variability in COVID-19 cases. Based on these findings, we further explore results from the full spatio-temporal model.

Method	WAIC	p_{WAIC}	PP Deviance
Full Spatiotemporal	**187,849.30**	612.45	**68,177.60**
Global Contact	188,562.30	334.83	69,360.68
No Contact	188,680.40	328.88	69,535.07

Table 2: Model fit (WAIC), complexity (p_{WAIC}), and posterior predictive (ppd) deviance results from each of the competing methods.

	Relative Risk	
Variable	Mean	95% CrI
Population[a]	1.34	(1.24, 1.44)
Median Household Income (USD)[a]	0.94	(0.88, 0.99)
% Poverty[b]	0.88	(0.66, 1.15)
% White, Non-Hispanic[b]	0.89	(0.81, 0.99)
% Some College Education[b]	0.95	(0.67, 1.31)
% $\geq$ 65 Years[b]	0.77	(0.58, 0.98)

Table 3: Regression parameter posterior inference (posterior means and quantile-based 95% credible intervals) from the full spatio-temporal model. (a) 10,000 increase; (b) 10% increase.

In Table 3, we display posterior inference for the non-contact regression parameters (β_j), exponentiated for a relative risk interpretation. Results suggest that towns with increased population have more cases on average, and that towns with higher median household income and a larger proportion of residents that are non-Hispanic White and $\geq$ 65 years have fewer cases. None of the other included spatially-varying covariates had 95% credible intervals that excluded one.

In Figure 11, we show the estimated global lag parameters, $\boldsymbol{\theta}$, and 95% credible intervals across the different lag days from the full spatio-temporal version of the model. The figure suggests that increased contact on prior days 3-7 is associated with increased current day COVID-19 cases.

2.5 Transmission model predicted cases and estimated infections for each town

Figure 12 through 33 show contact, projected infections, cases, and cumulative incidence for each of the 169 towns in Connecticut.

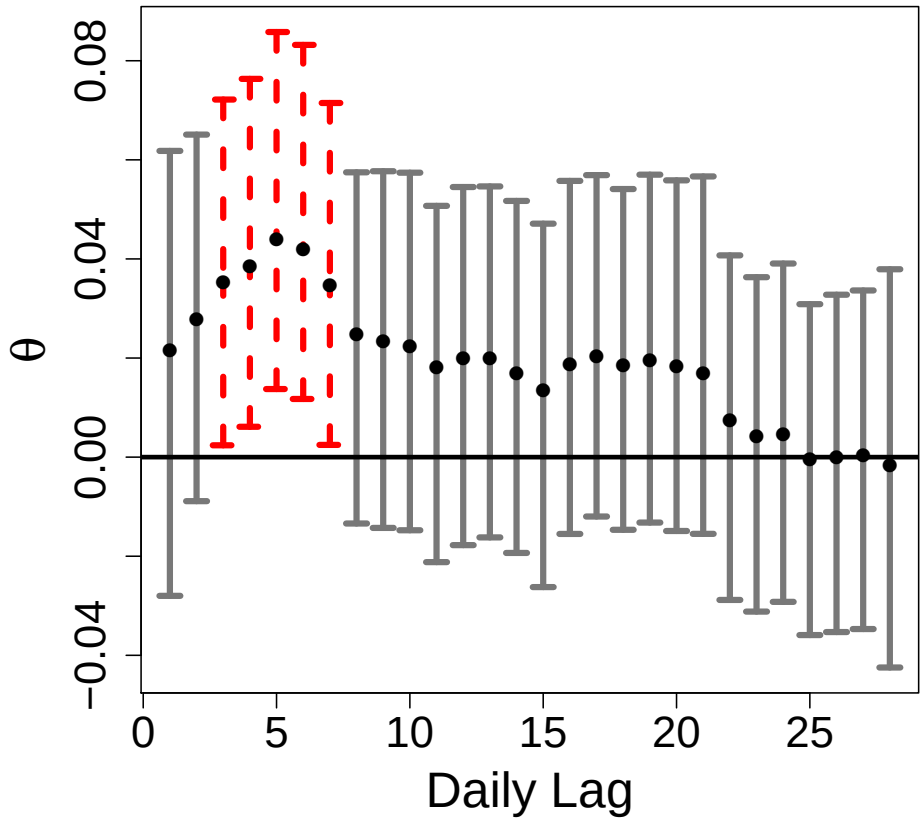

Figure 11: Estimated daily lag regression parameters (and 95% credible intervals) describing associations between contact and current day COVID-19 cases. Red bars indicate that the credible intervals exclude zero.

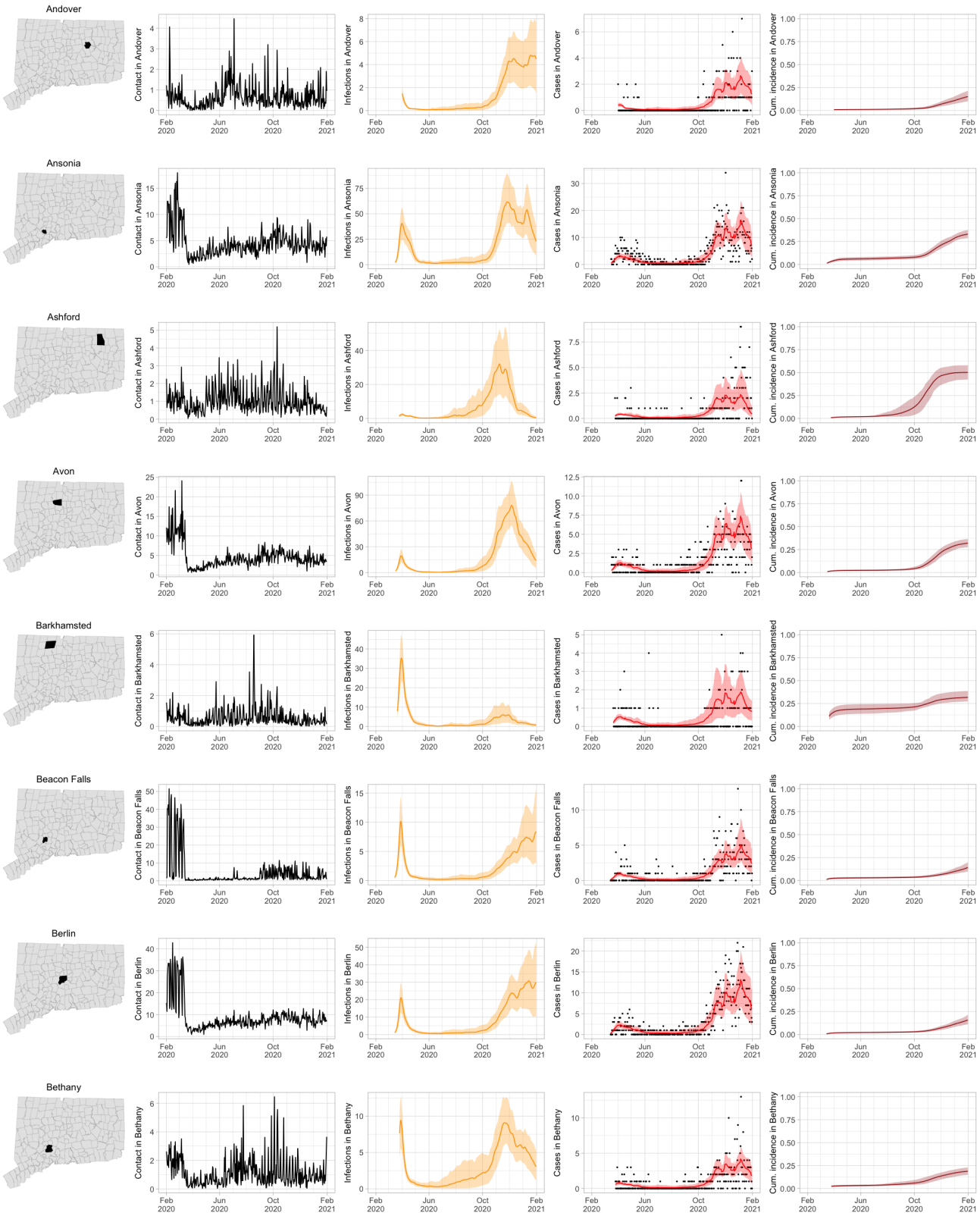

Figure 12: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

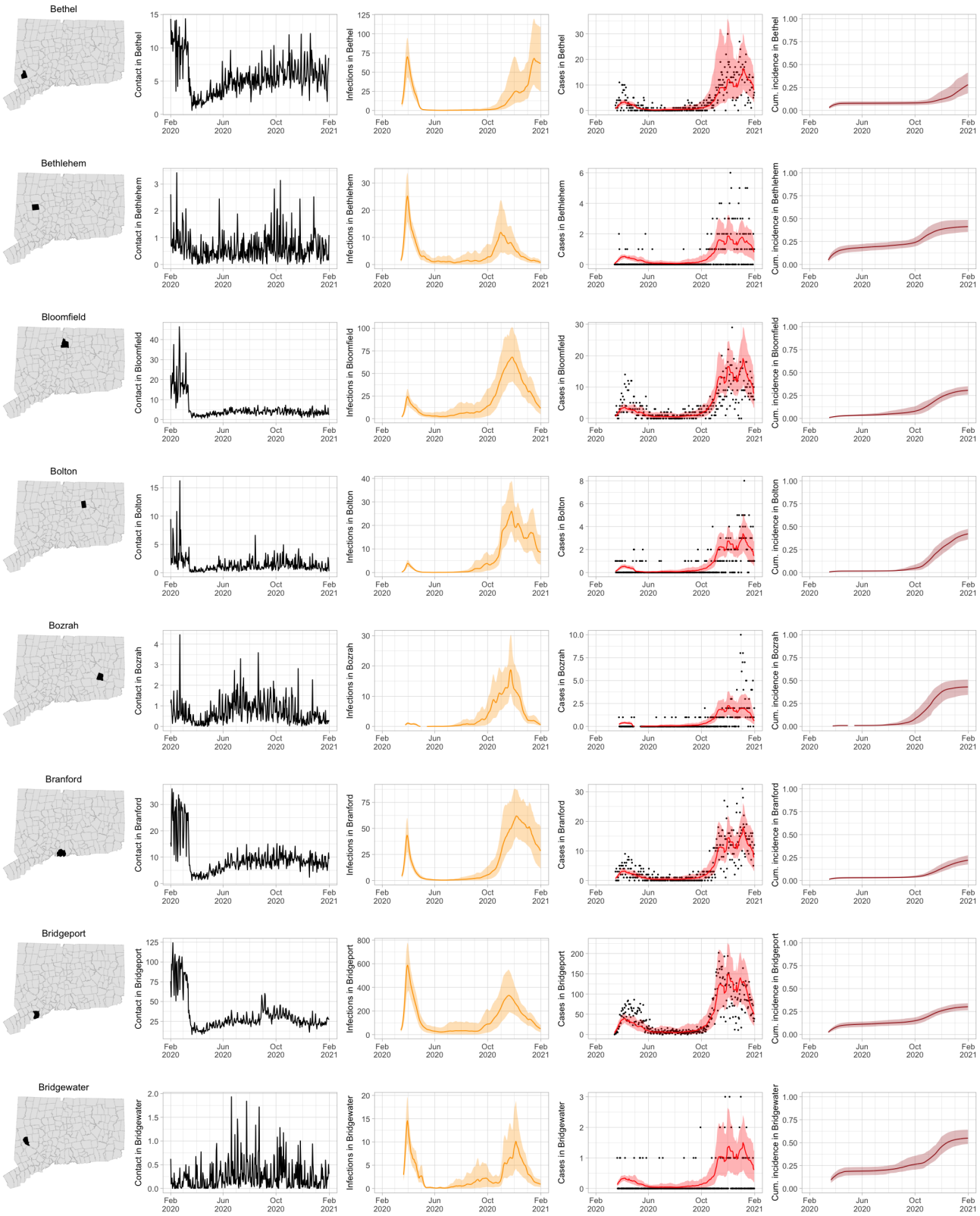

Figure 13: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

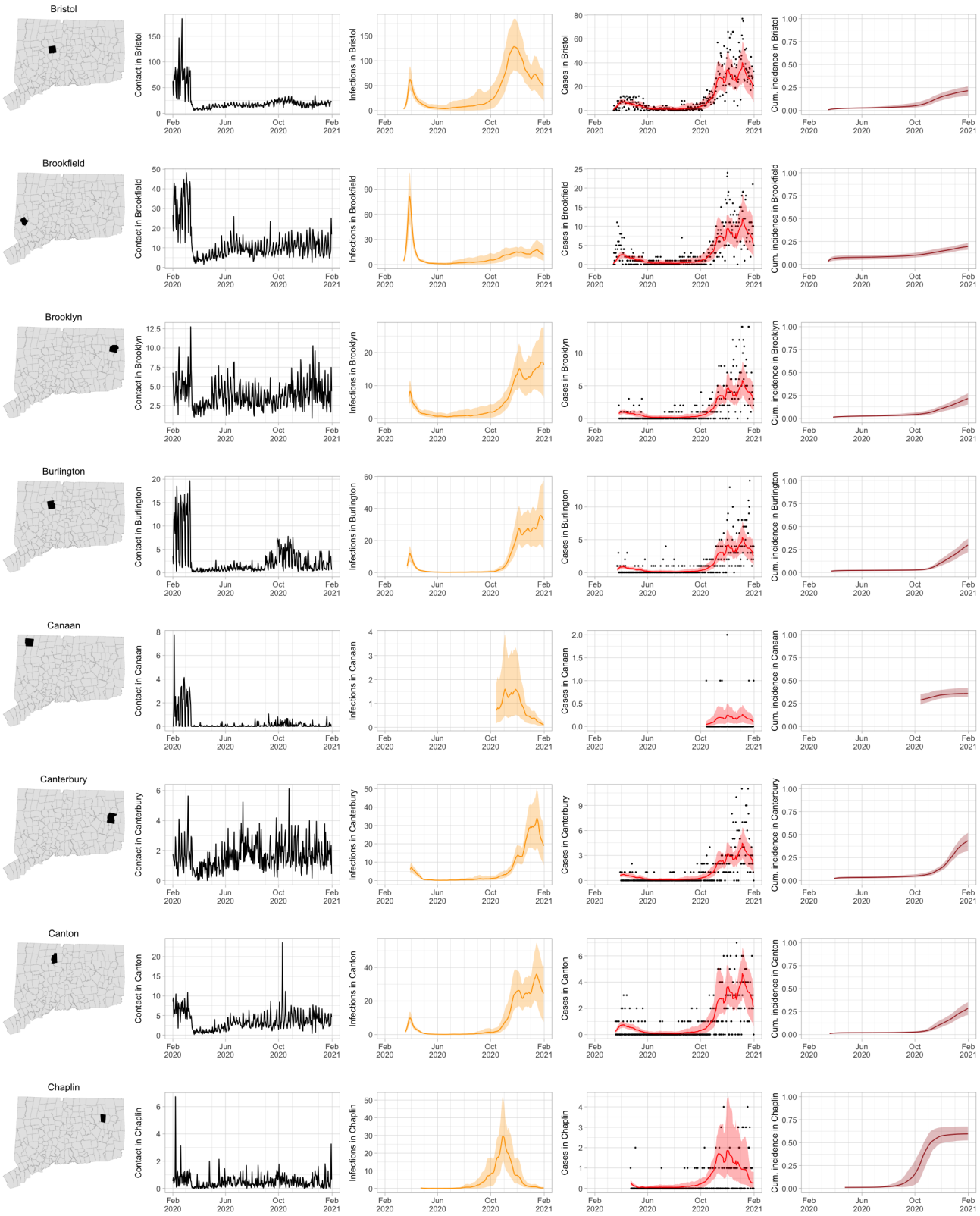

Figure 14: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

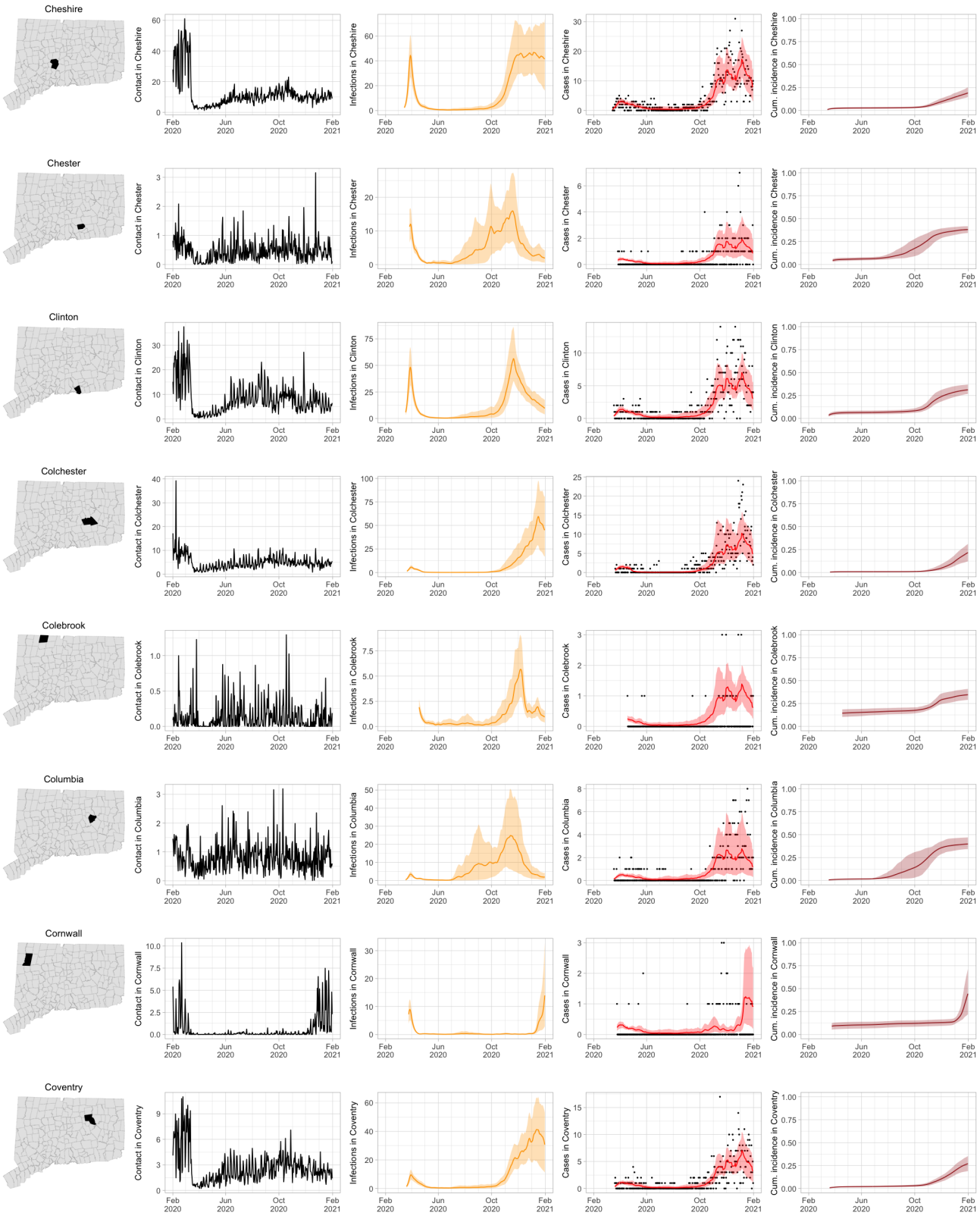

Figure 15: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

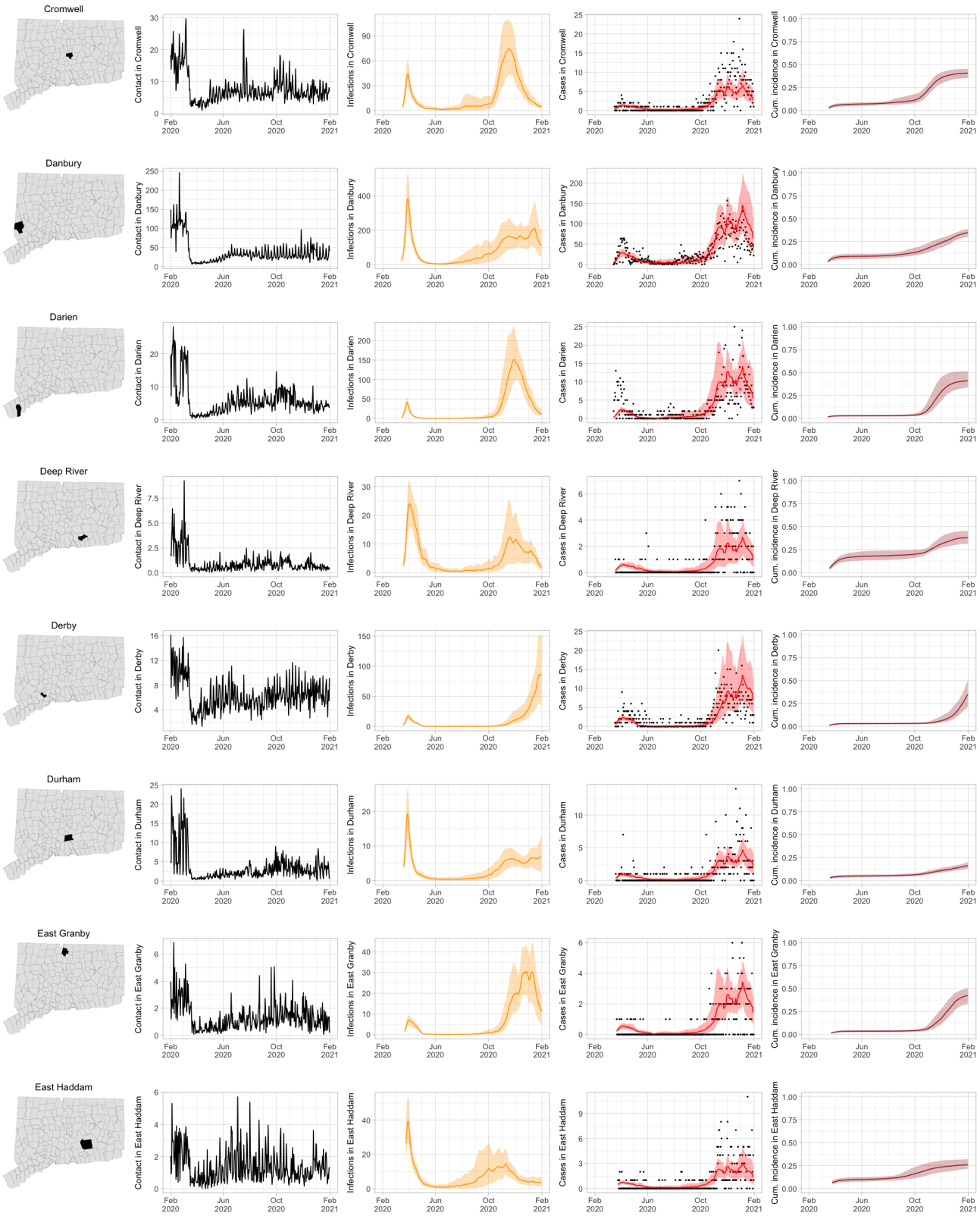

Figure 16: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

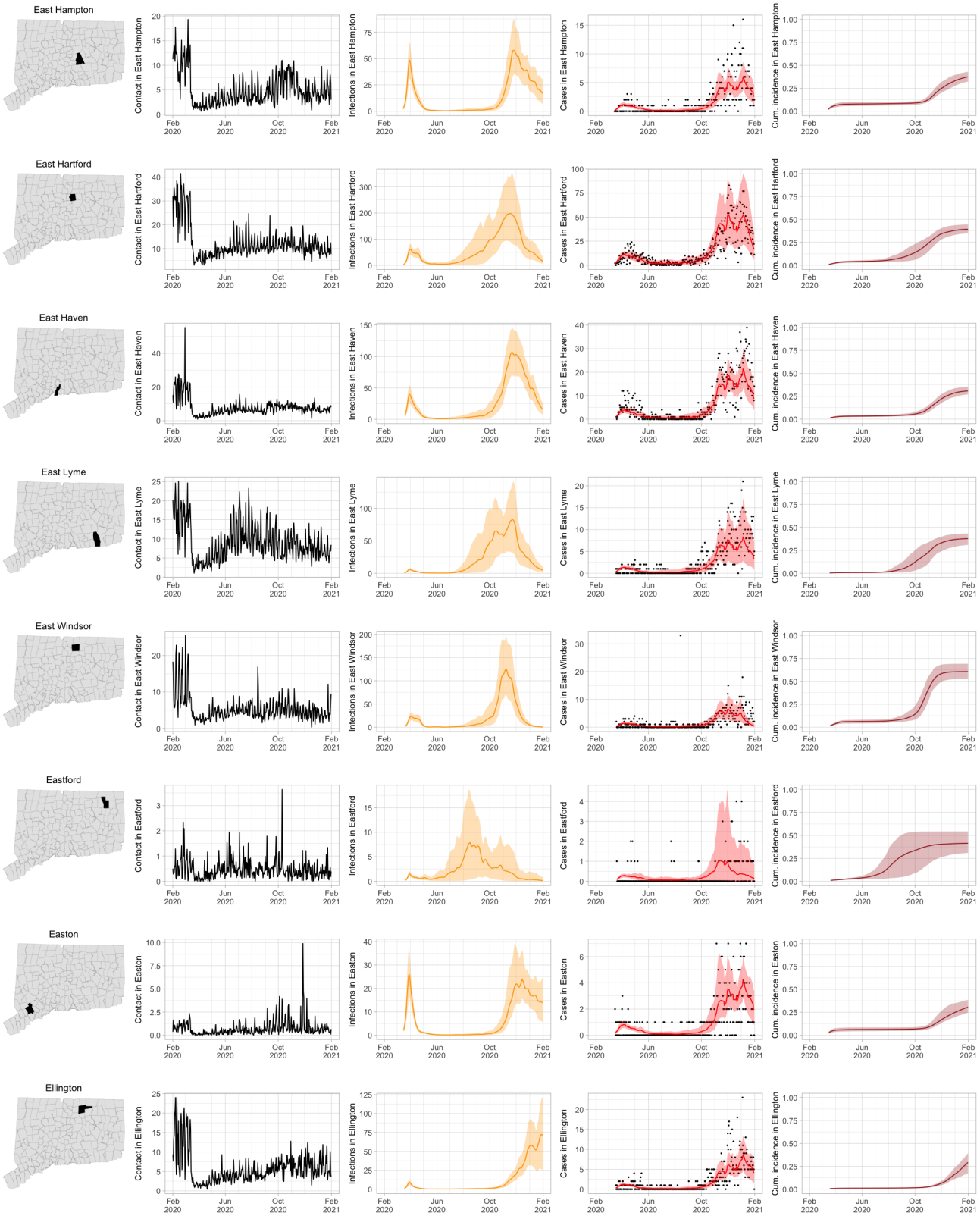

Figure 17: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

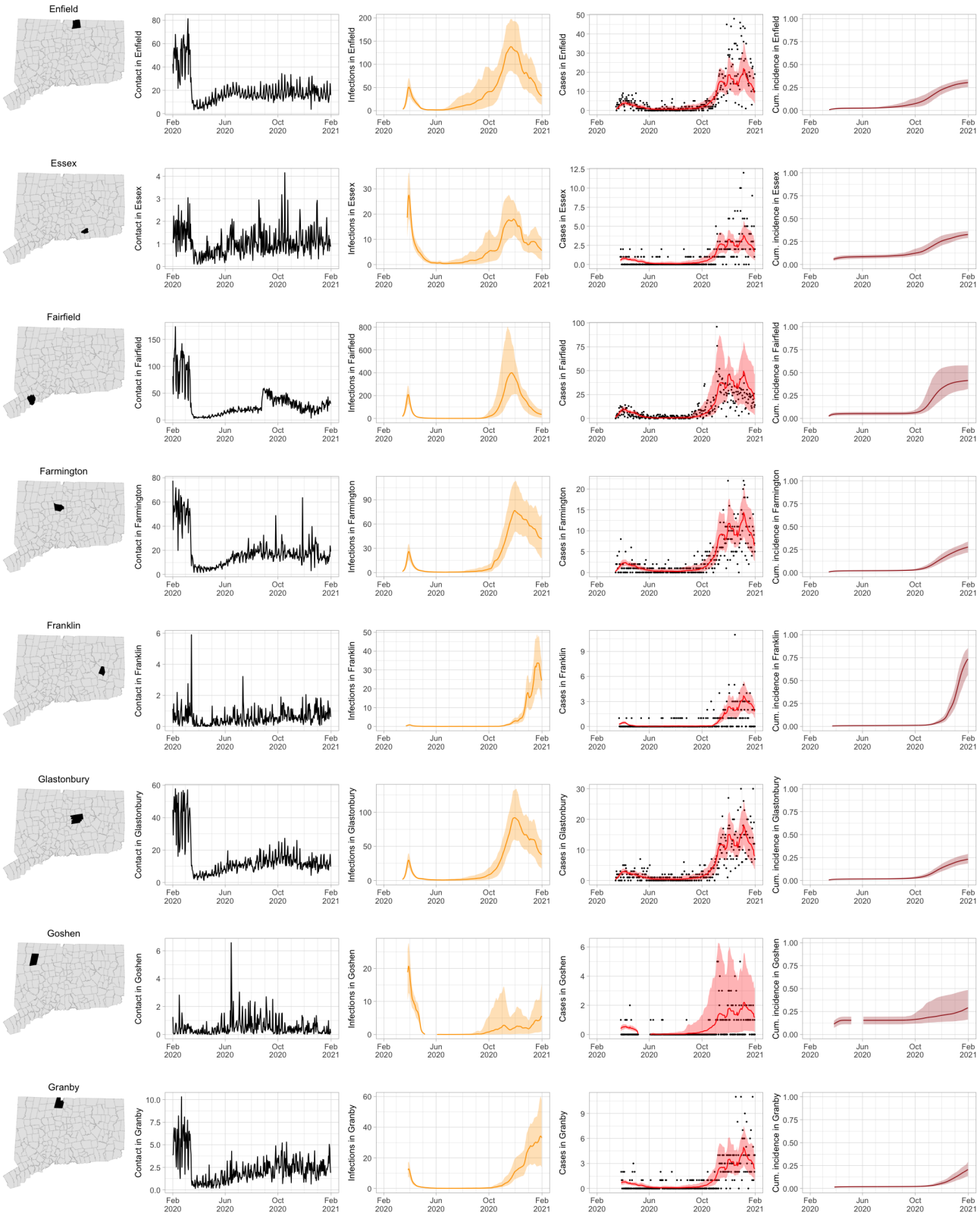

Figure 18: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

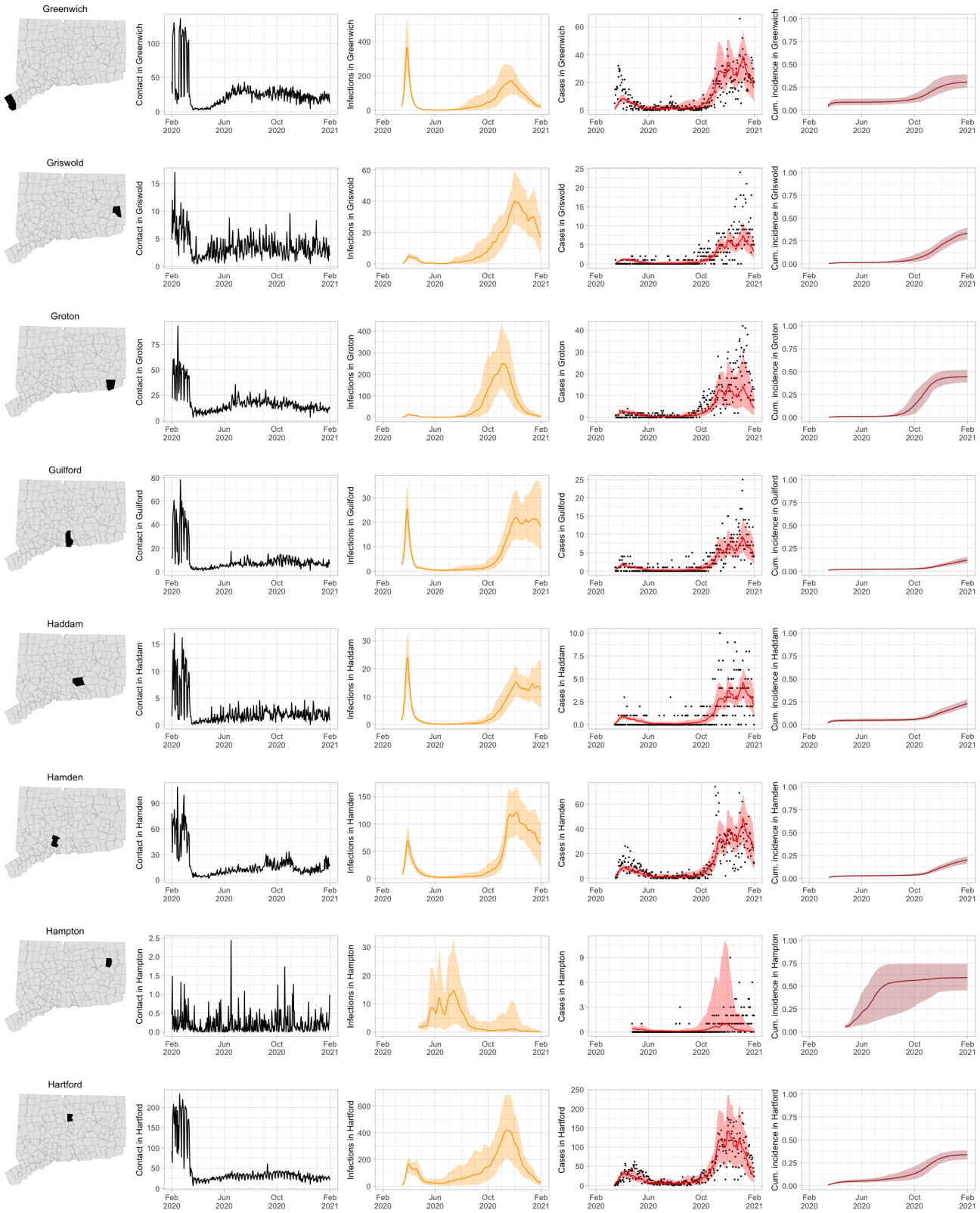

Figure 19: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

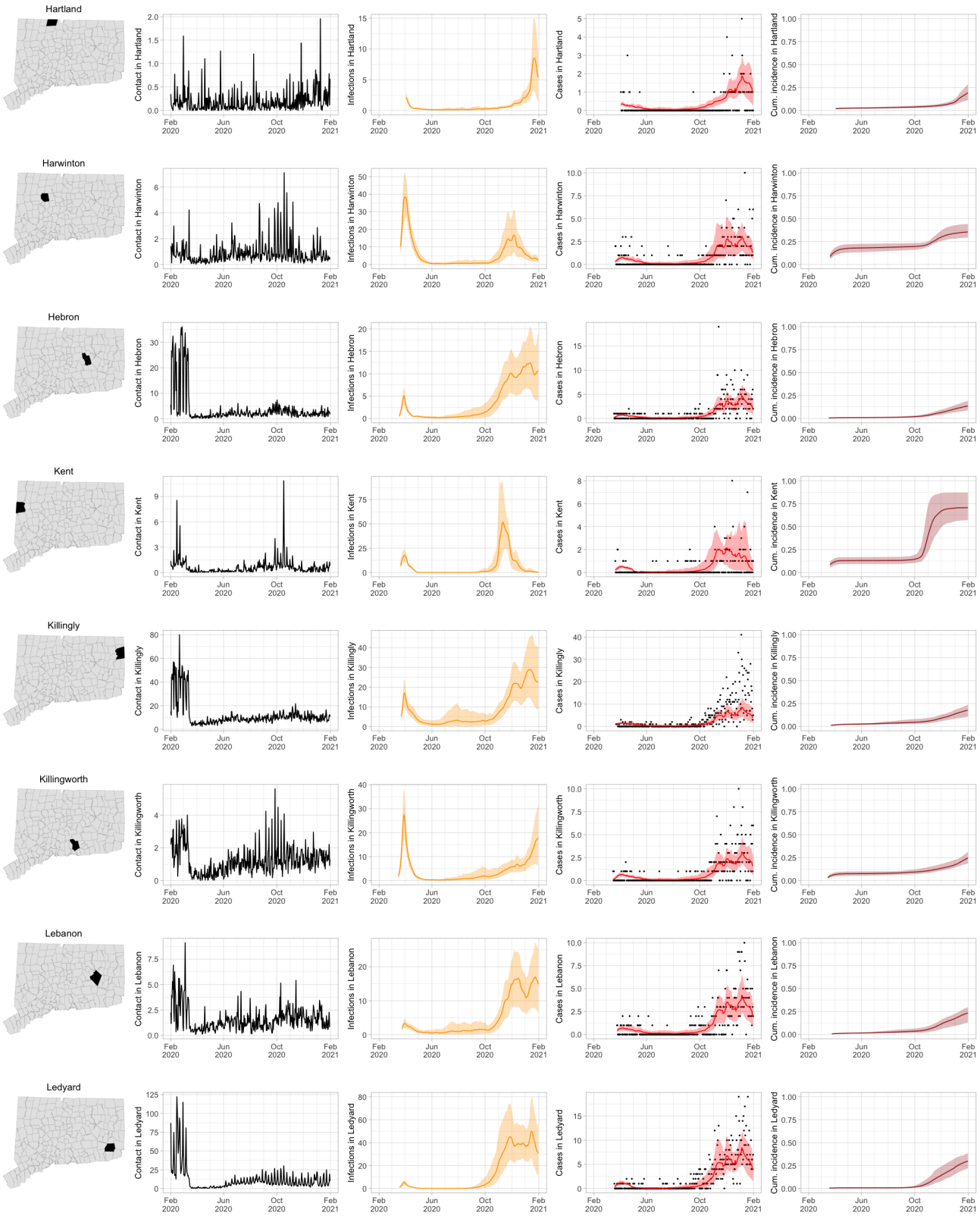

Figure 20: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

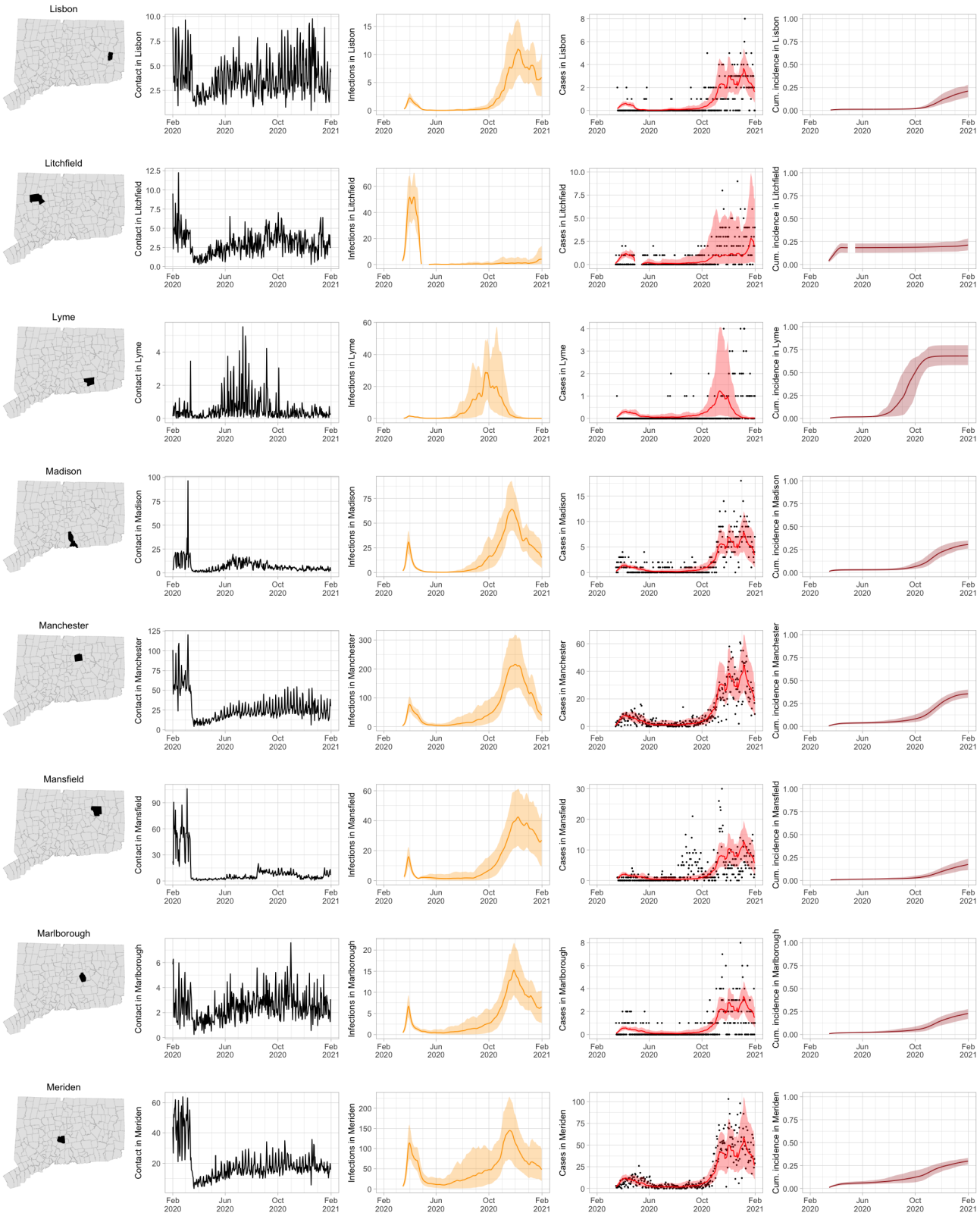

Figure 21: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

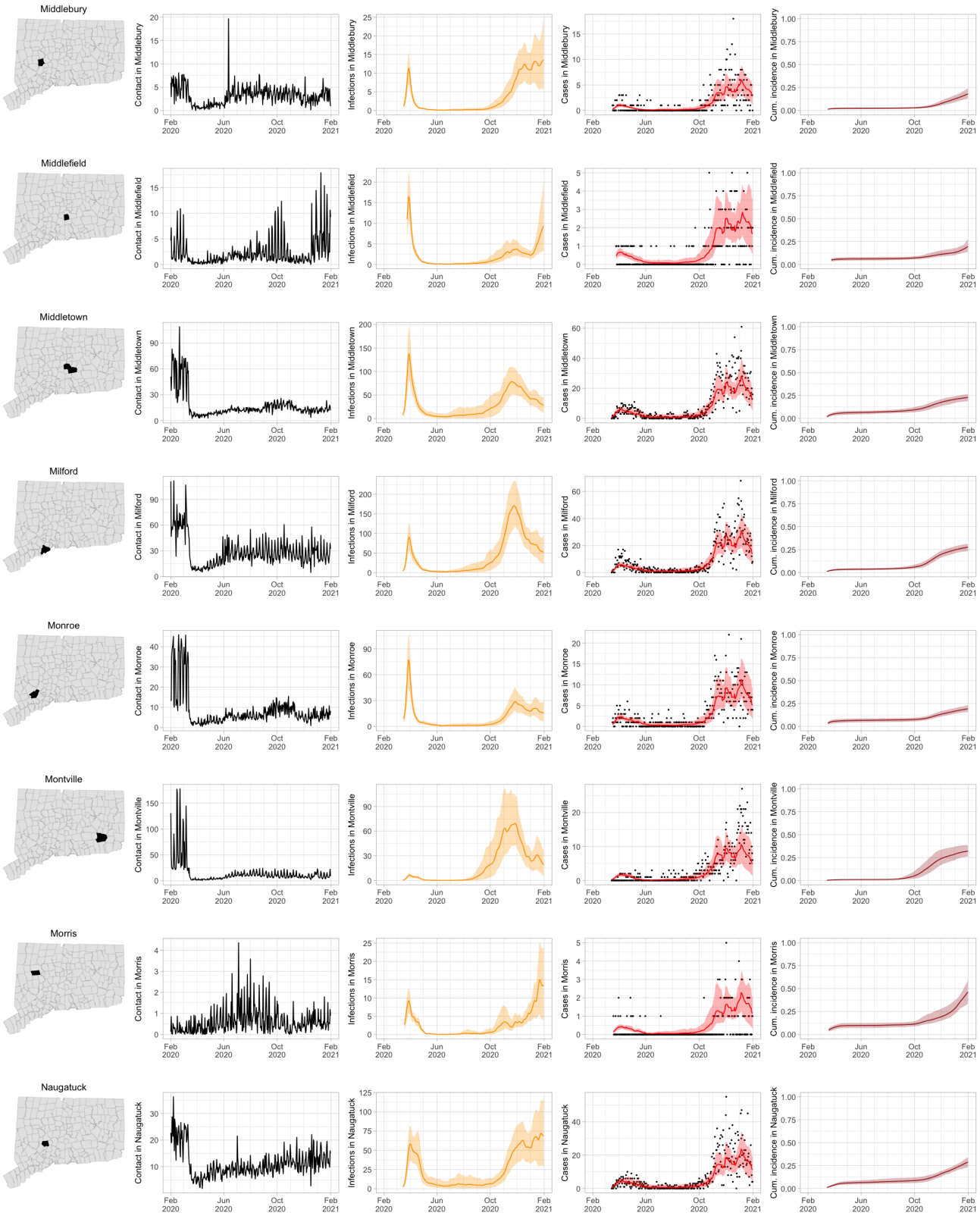

Figure 22: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

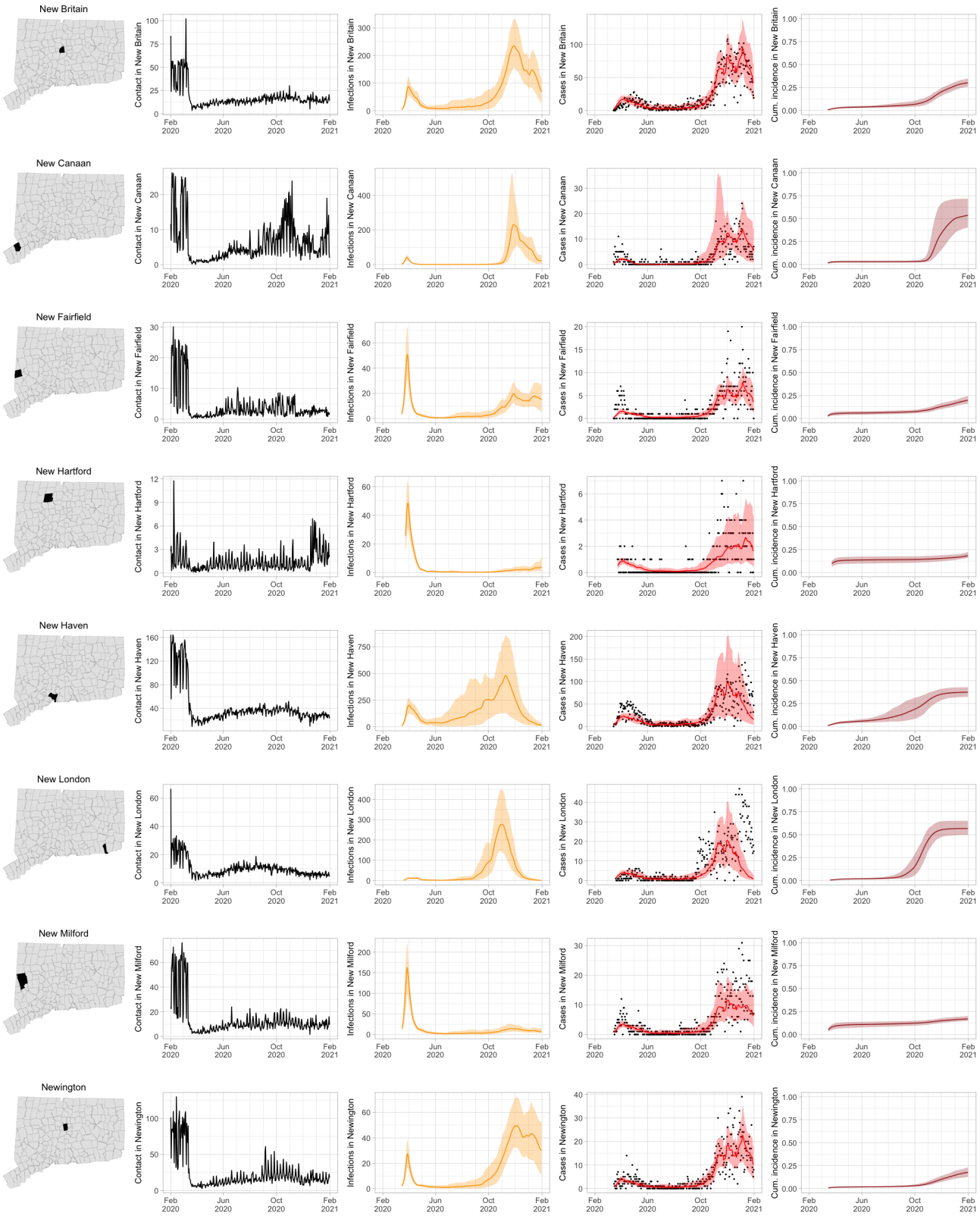

Figure 23: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

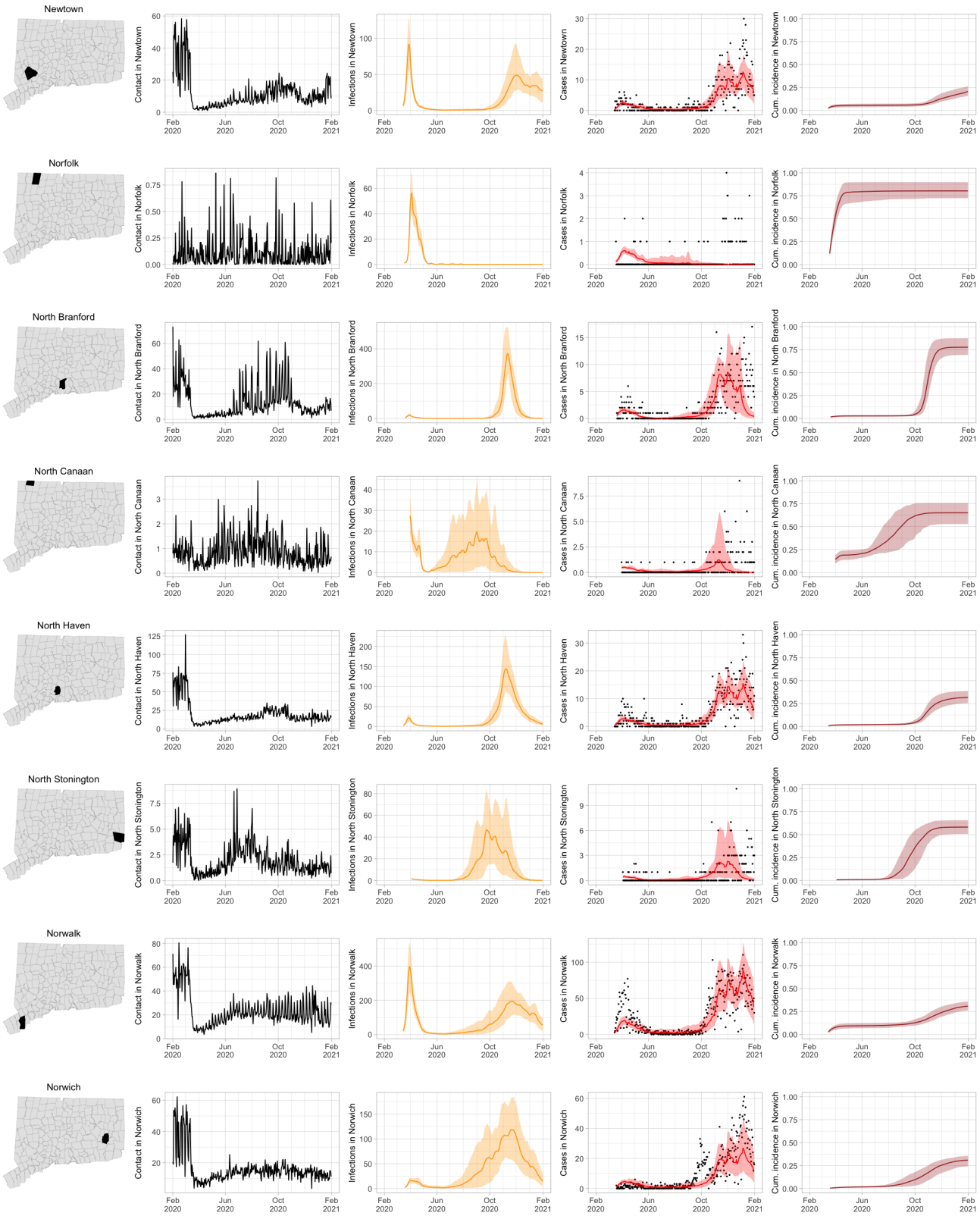

Figure 24: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

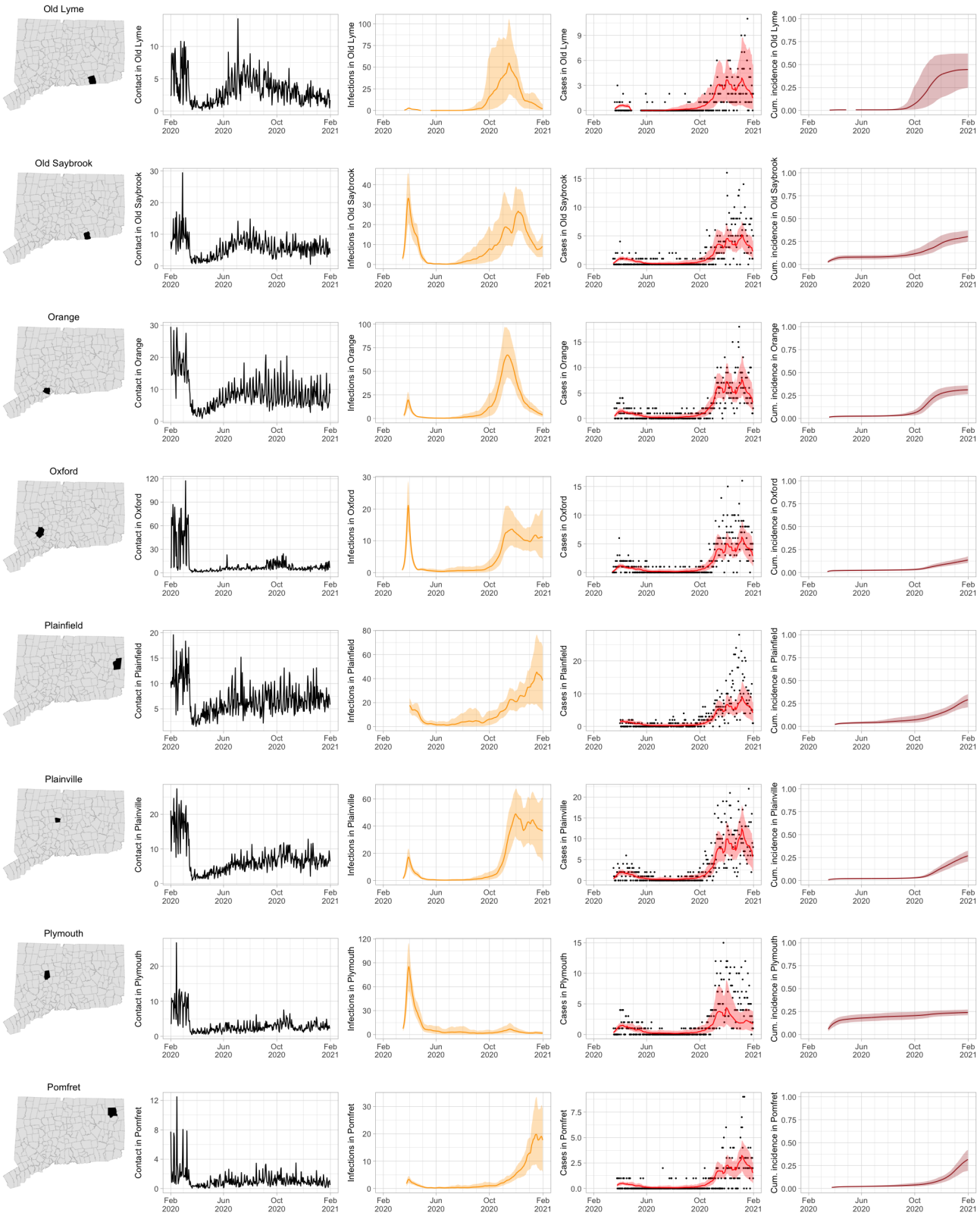

Figure 25: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

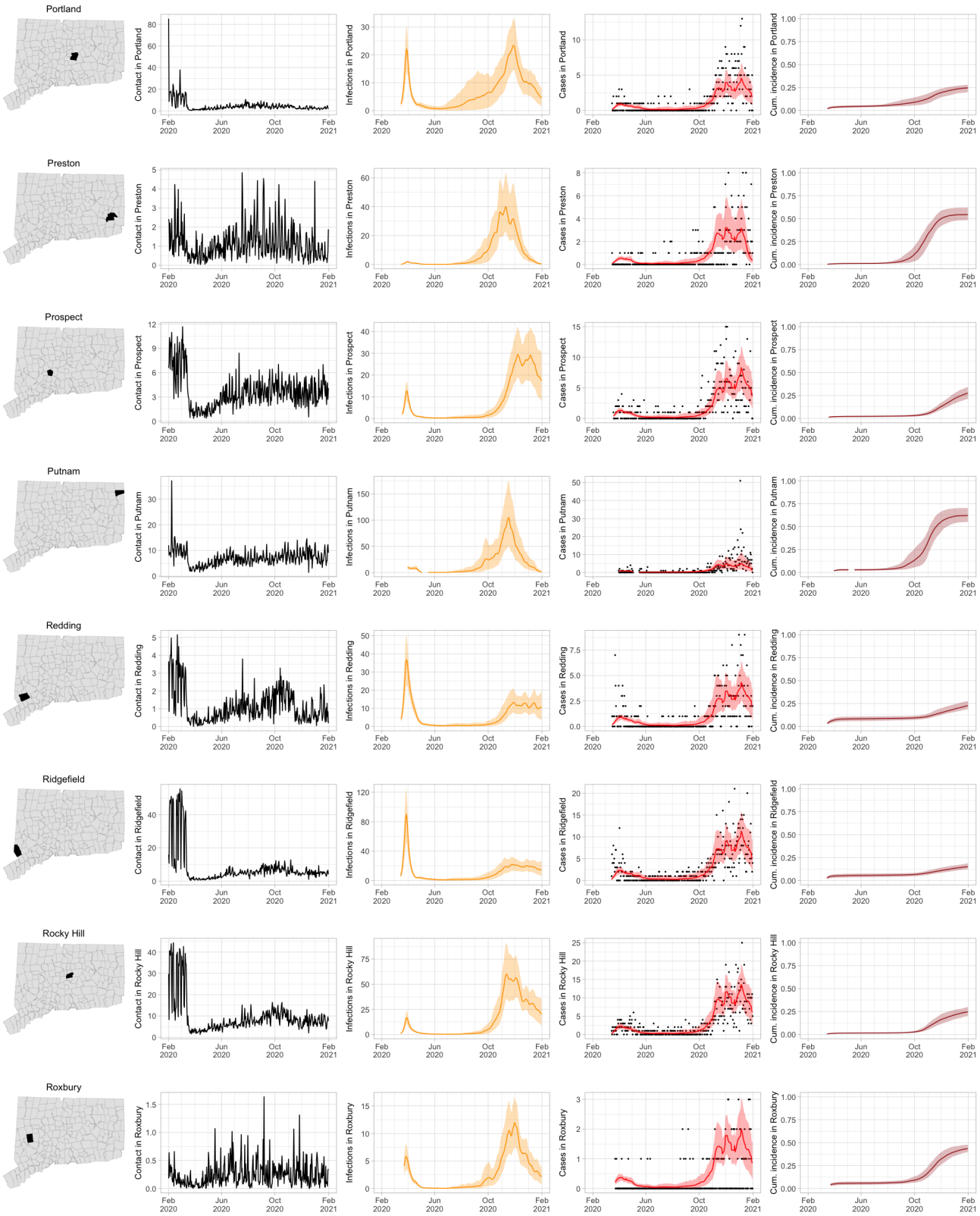

Figure 26: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

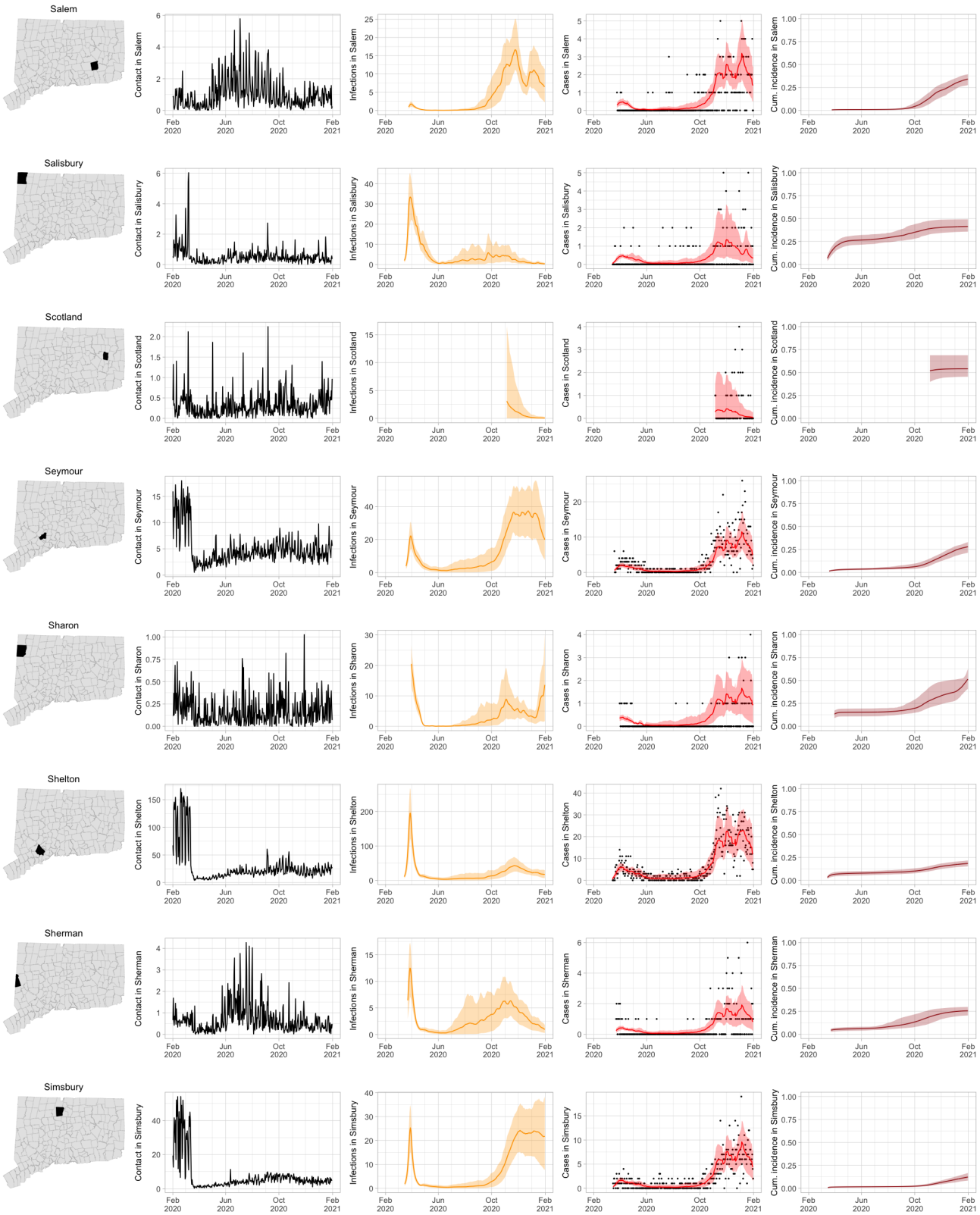

Figure 27: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

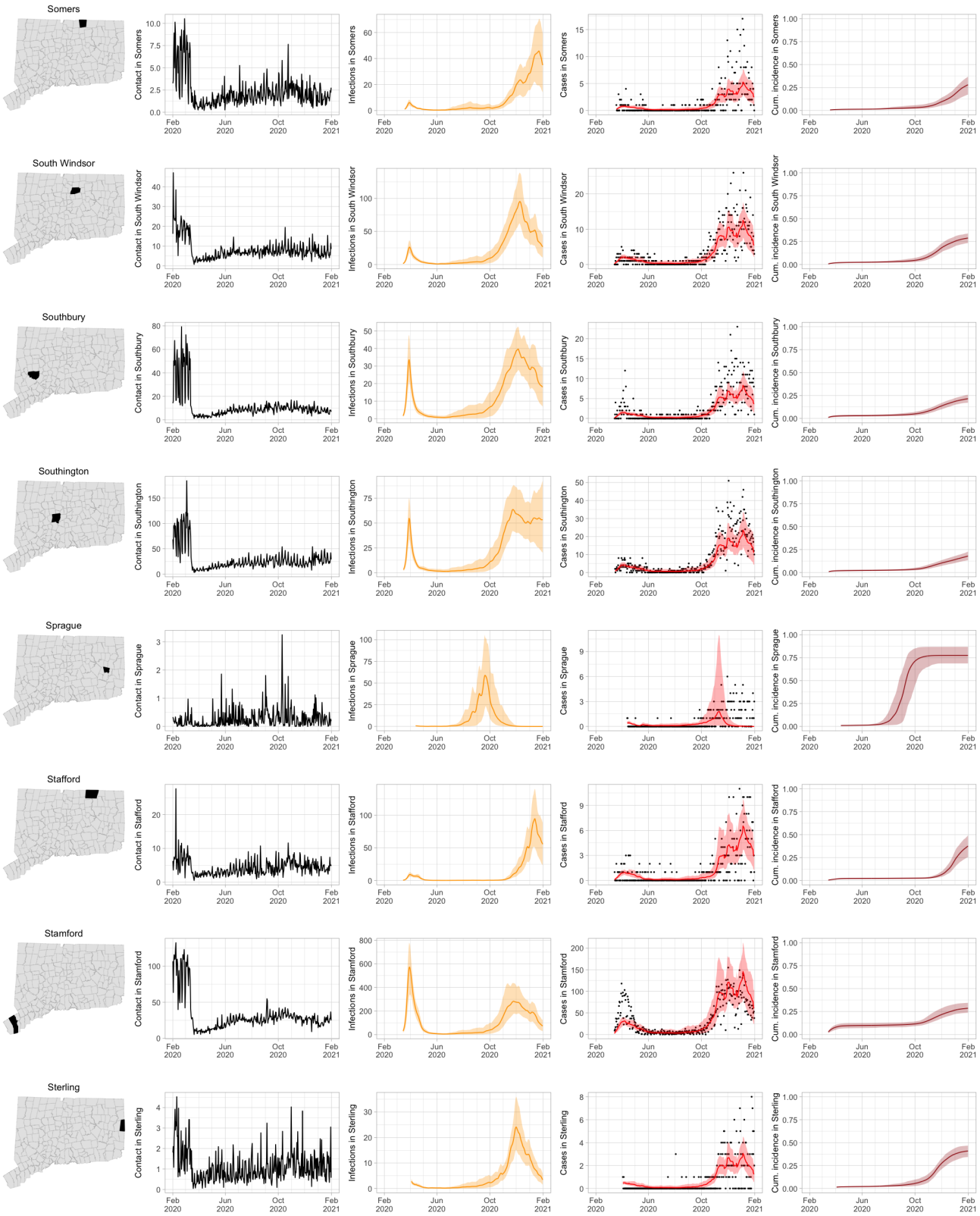

Figure 28: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

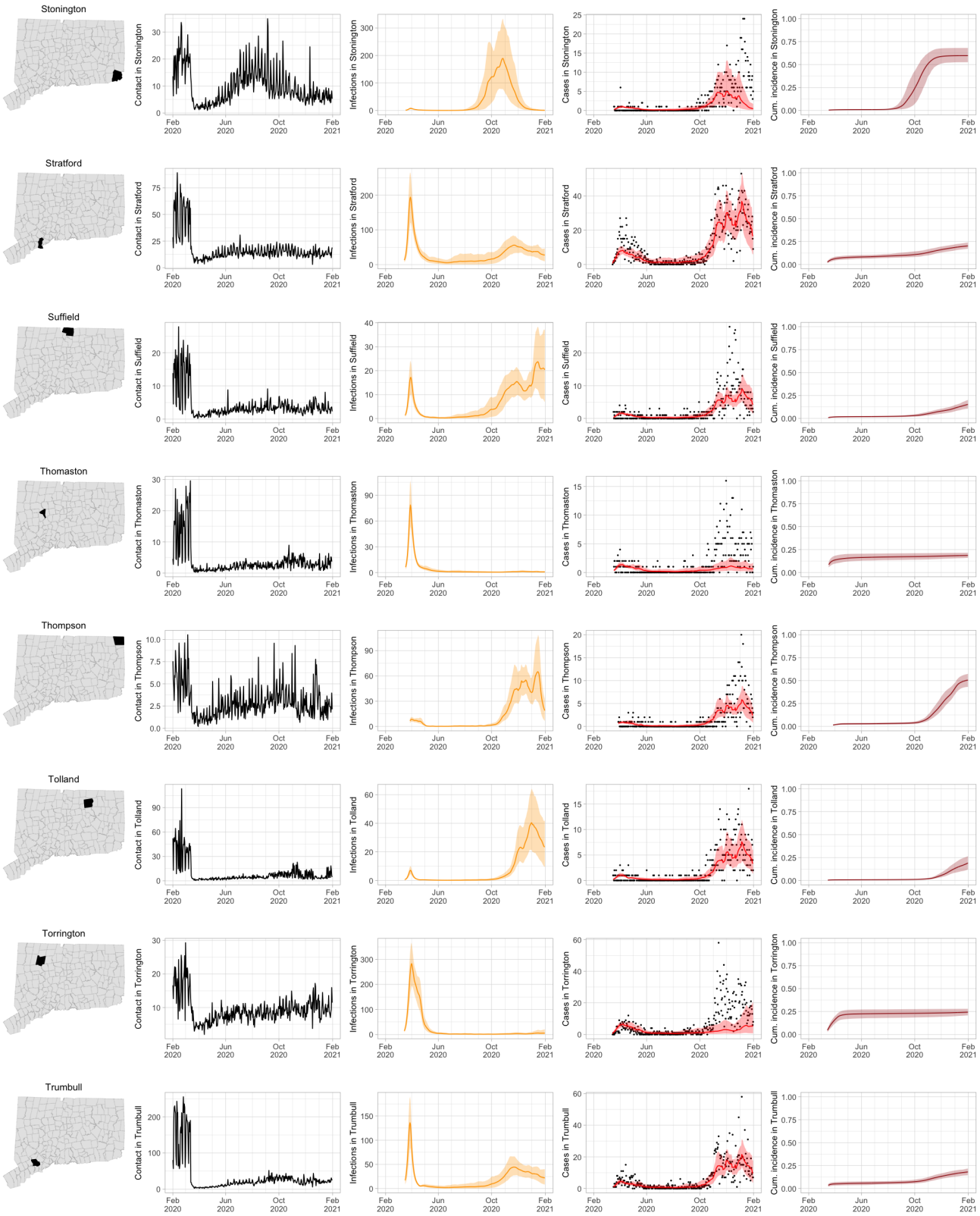

Figure 29: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

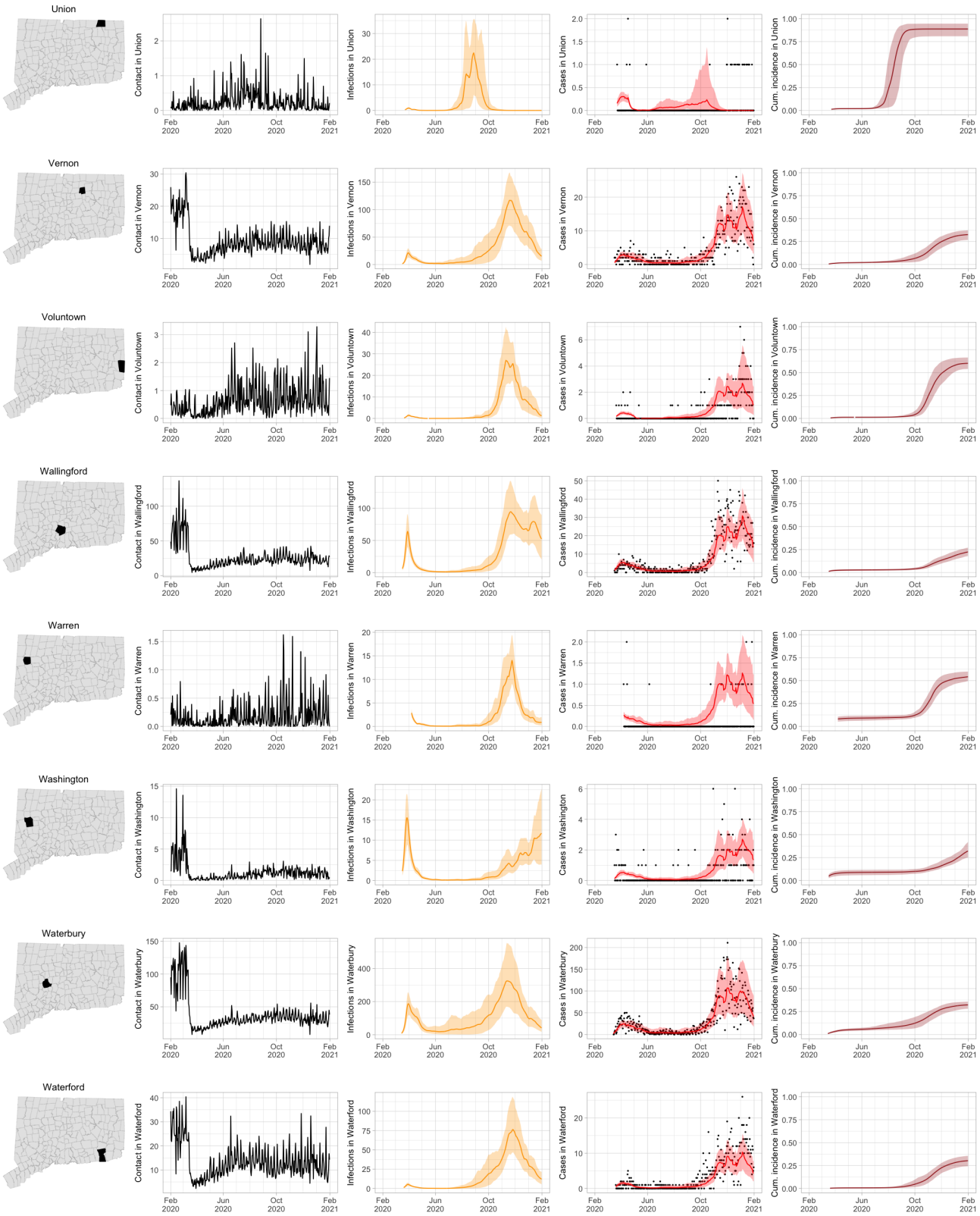

Figure 30: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

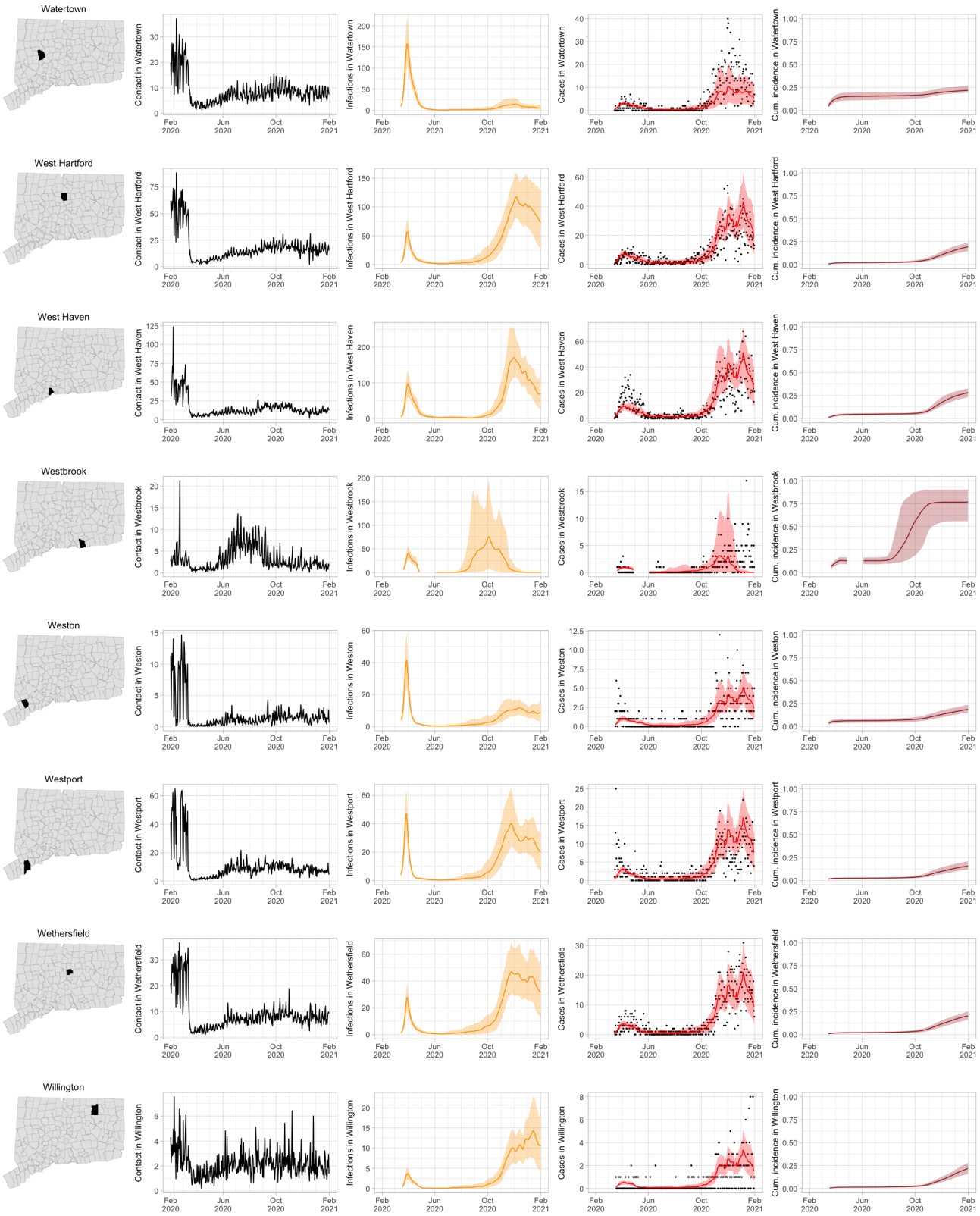

Figure 31: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

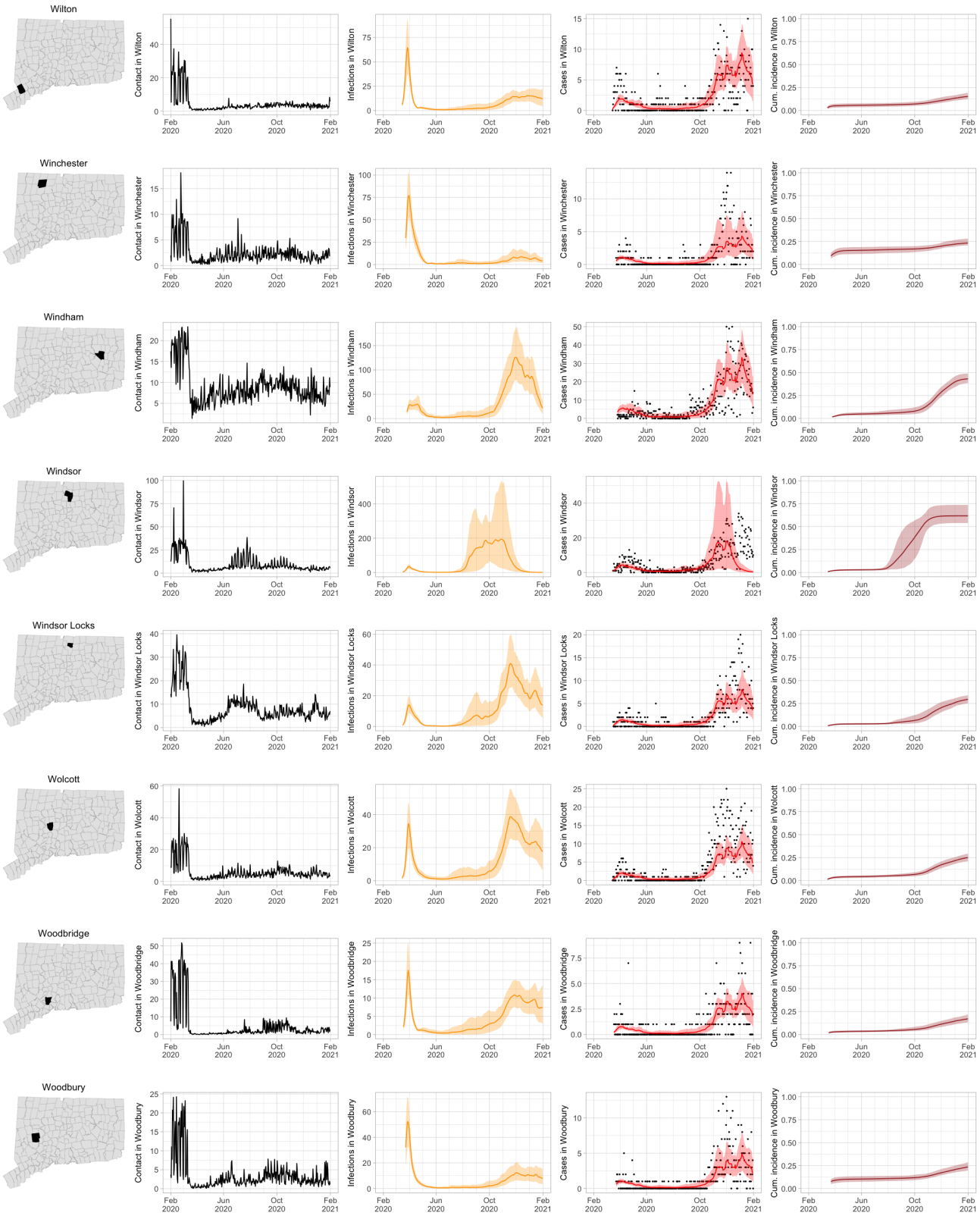

Figure 32: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

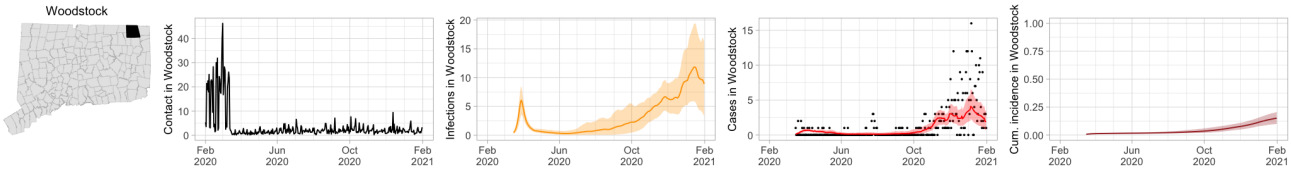

Figure 33: Contact for Connecticut towns and fitted SEIR model predictions with 95% uncertainty intervals.

References

[1] Glen Van Brummelen. *Heavenly mathematics: The forgotten art of spherical trigonometry*. Princeton University Press, 2012.

[2] Change to the list of reportable diseases, emergency illnesses and health conditions and the list of reportable laboratory findings. *Connecticut Epidemiologist*, 2020. URL `https://portal.ct.gov/-/media/DPH/EEIP/CTEPI/Vol40_No2.pdf`.

[3] Centers for Disease Control and Prevention. Coronavirus disease 2019 (covid-19) 2020 interim case definition, approved august 5, 2020, 2020. URL `https://wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/08/05/`.

[4] Olga Morozova, Zehang Richard Li, and Forrest W Crawford. A model for COVID-19 transmission in Connecticut. *medRxiv*, 2021. doi: https://doi.org/10.1101/2020.06.12.20126391. URL `https://www.medrxiv.org/content/10.1101/2020.06.12.20126391v2`.

[5] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

[6] Karline Soetaert, Thomas Petzoldt, and R. Woodrow Setzer. Solving differential equations in R: Package deSolve. *Journal of Statistical Software*, 33(9):1–25, 2010.

[7] Tegan K Boehmer, Jourdan DeVies, Elise Caruso, Katharina L van Santen, Shichao Tang, Carla L Black, Kathleen P Hartnett, Aaron Kite-Powell, Stephanie Dietz, Matthew Lozier, et al. Changing age distribution of the COVID-19 pandemic — United States, May–August 2020. *Morbidity and Mortality Weekly Report*, 69(39):1404, 2020.

[8] Daniel P Oran and Eric J Topol. Prevalence of asymptomatic SARS-CoV-2 infection: A narrative review. *Annals of Internal Medicine*, 2020.

[9] Diana Buitrago-Garcia, Dianne Egli-Gany, Michel J Counotte, Stefanie Hossmann, Hira Imeri, Aziz Mert Ipekci, Georgia Salanti, and Nicola Low. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS medicine*,

17(9):e1003346, 2020.

[10] Centers for Disease Control and Prevention. COVID-19 Pandemic Planning Scenarios. https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html, 2021. Accessed: 2021-01-18.

[11] Stephen M Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H Grad, and Marc Lipsitch. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, 368(6493):860–868, 2020.

[12] Henrik Salje, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, Yann Le Strat, Justin Lessler, Daniel Levy Bruhl, Arnaud Fontanet, Lulla Opatowski, Pierre-Yves Boelle, and Simon Cauchemez. Estimating the burden of SARS-CoV-2 in France. *Science*, 369(6500):208–211, 2020.

[13] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick GT Walker, Han Fu, Amy Dighe, Jamie T Griffin, Marc Baguelin, Sangeeta Bhatia, Anne Boonyasiri, Adhiratha andd Cori, Zulma Cucunubá, Rich FitzJohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A Donnelly, Azra C Ghani, and Neil M Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6):669–677, 2020.

[14] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, 2020.

[15] Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini Jr, Stefano Merler, Alex Pentland, Alessandro Vespignani, Esteban Moro, and Yamir Moreno. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour*, 4(9):964–971, 2020.

[16] Andrew William Byrne, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, Kirsty O'Brien, Patrick Wall, Kieran A Walsh, and Simon J More. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open*, 10(8), 2020. ISSN 2044-6055. doi: 10.1136/bmjopen-2020-039856. URL https://bmjopen.bmj.com/content/10/8/e039856.

[17] Wycliffe E Wei, Zongbin Li, Calvin J Chiew, Sarah E Yong, Matthias P Toh, and Vernon J Lee. Presymp-

tomatic transmission of SARS-CoV-2 – Singapore, January 23–March 16, 2020. *Morbidity and Mortality Weekly Report*, 69(14):411, 2020.

[18] Matthew Biggerstaff, Benjamin J Cowling, Zulma M Cucunubá, Linh Dinh, Neil M Ferguson, Huizhi Gao, Verity Hill, Natsuko Imai, Michael A Johansson, Sarah Kada, et al. Early insights from statistical and mathematical modeling of key epidemiologic parameters of COVID-19. *Emerging infectious diseases*, 26 (11), 2020.

[19] Conor McAloon, Áine Collins, Kevin Hunt, Ann Barber, Andrew W Byrne, Francis Butler, Miriam Casey, John Griffin, Elizabeth Lane, David McEvoy, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ open*, 10(8):e039652, 2020.

[20] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 172 (9):577–582, 2020.

[21] Jantien A Backer, Don Klinkenberg, and Jacco Wallinga. Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, china, 20–28 january 2020. *Eurosurveillance*, 25(5), 2020.

[22] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy S.M. Leung, Eric H.Y. Lau, Jessica Y. Wong, Xuesen Xing, Nijuan Xiang, Yang Wu, Chao Li, Qi Chen, Dan Li, Tian Liu, Jing Zhao, Man Liu, Wenxiao Tu, Chuding Chen, Lianmei Jin, Rui Yang, Qi Wang, Suhua Zhou, Rui Wang, Hui Liu, Yinbo Luo, Yuan Liu, Ge Shao, Huan Li, Zhongfa Tao, Yang Yang, Zhiqiang Deng, Boxi Liu, Zhitao Ma, Yanping Zhang, Guoqing Shi, Tommy T.Y. Lam, Joseph T. Wu, George F. Gao, Benjamin J. Cowling, Bo Yang, Gabriel M. Leung, and Zijian Feng. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, 2020.

[23] Jing Qin, Chong You, Qiushi Lin, Taojun Hu, Shicheng Yu, and Xiao-Hua Zhou. Estimation of incubation period distribution of COVID-19 using disease onset forward time: A novel cross-sectional and forward follow-up study. *Science Advances*, 6(33), 2020. doi: $10.1126/\mathrm{sciadv.abc1202}$. URL `https://advances.sciencemag.org/content/6/33/eabc1202`.

[24] Muge Cevik, Matthew Tate, Ollie Lloyd, Alberto Enrico Maraolo, Jenna Schafers, and Antonia Ho. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, 2020.

[25] Kieran A Walsh, Karen Jordan, Barbara Clyne, Daniela Rohde, Linda Drummond, Paula Byrne, Susan

Ahern, Paul G Carty, Kirsty K O'Brien, Eamon O'Murchu, et al. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *Journal of Infection*, 2020.

[26] Xueting Qiu, Ali Ihsan Nergiz, Alberto Enrico Maraolo, Isaac I. Bogoch, Nicola Low, and Muge Cevik. Differences in secondary attack rates based on symptom status of index case(s)– a living systematic review. *medRxiv*, 2021. doi: 10.1101/2020.09.01.20135194. URL `https://www.medrxiv.org/content/early/2021/01/16/2020.09.01.20135194`.

[27] Rongrong Yang, Xien Gui, and Yong Xiong. Comparison of clinical characteristics of patients with asymptomatic vs symptomatic coronavirus disease 2019 in Wuhan, China. *JAMA Network Open*, 3(5): e2010182–e2010182, 2020.

[28] Kieran A Walsh, Susan Spillane, Laura Comber, Karen Cardwell, Patricia Harrington, Jeff Connell, Conor Teljeur, Natasha Broderick, Cillian F de Gascun, Susan M Smith, et al. The duration of infectiousness of individuals infected with SARS-CoV-2. *Journal of Infection*, 2020.

[29] Roman Wölfel, Victor M. Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A. Müller, Daniela Niemeyer, Terry C. Jones, Patrick Vollmar, Camilla Rothe, Michael Hoelscher, Tobias Bleicker, Sebastian Brünink, Julia Schneider, Rosina Ehmann, Katrin Zwirglmaier, Christian Drosten, and Clemens Wendtner. Virological assessment of hospitalized patients with COVID-2019. *Nature*, 581(7809): 465–469, 2020.

[30] David Mc Evoy, Conor G McAloon, Aine B Collins, Kevin Hunt, Francis Butler, Andrew W Byrne, Miriam Casey, Ann Barber, John M Griffin, Elizabeth A Lane, et al. The relative infectiousness of asymptomatic SARS-CoV-2 infected persons compared with symptomatic individuals: A rapid scoping review. *medRxiv*, 2020. doi: https://doi.org/10.1101/2020.07.30.20165084. URL `https://www.medrxiv.org/content/10.1101/2020.07.30.20165084v1`.

[31] Daihai He, Shi Zhao, Qianying Lin, Zian Zhuang, Peihua Cao, Maggie H Wang, and Lin Yang. The relative transmissibility of asymptomatic COVID-19 infections among close contacts. *International Journal of Infectious Diseases*, 2020.

[32] Centers for Disease Control and Prevention. Discontinuation of isolation for persons with COVID-19 not in healthcare settings. interim guidance. `https://www.cdc.gov/coronavirus/2019-ncov/hcp/disposition-in-home-patients.html`, 2021. Accessed: 2021-01-18.

[33] Shikha Garg, Lindsay Kim, Michael Whitaker, Alissa O'Halloran, Charisse Cummings, Rachel Holstein, Mila Prill, Shua J Chai, Pam D Kirley, Nisha B Alden, Breanna Kawasaki, Kimberly Yousey-Hindes, Linda Niccolai, Evan J Anderson, Kyle P Openo, Andrew Weigel, Maya L Monroe, Patricia Ryan, Justin Hen-

derson, Sue Kim, Kathy Como-Sabetti, Ruth Lynfield, Daniel Sosin, Salina Torres, Alison Muse, Nancy M Bennett, Laurie Billing, Melissa Sutton, Nicole West, William Schaffner, H. Keipp Talbot, Aquino Clarissa, Andrea George, Alicia Budd, Lynnette Brammer, Gayle Langley, Aron J Hall, and Alicia Fry. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 – COVID-NET, 14 states, March 1–30, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69, 2020.

[34] Pablo N Perez-Guzman, Anna Daunt, Sujit Mukherjee, Peter Crook, Roberta Forlano, Mara D Kont, Alessandra Løchen, Michaela Vollmer, Paul Middleton, Rebekah Judge, , Chris Harlow, Anet Soubieres, Graham Cooke, Peter J White, Timothy B Hallett, Paul Aylin, Neil Ferguson, Katharina Hauck, Mark Thursz, and Shevanthi Nayagam. Clinical characteristics and predictors of outcomes of hospitalized patients with coronavirus disease 2019 in a multiethnic London national health service trust: a retrospective cohort study. *Clinical Infectious Diseases*, 2020.

[35] Connecticut Hospital Association, 2021. URL `https://cthosp.org/`.

[36] Connecticut State Department of Public Health. COVID-19 Data Resources, 2021. URL `https://data.ct.gov/stories/s/COVID-19-data/wa3g-tfvc/`.

[37] United States Census Bureau. *American Community Survey (ACS)*, 2020. URL `https://www.census.gov/programs-surveys/acs`.

[38] CHIMEData, 2020. URL `https://chimedata.org/`.

[39] Iain Murray, Ryan Adams, and David MacKay. Elliptical slice sampling. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 541–548, 2010.

[40] Brian G Leroux, Xingye Lei, and Norman Breslow. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer, 2000.

[41] Sumio Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594, 2010.

[42] Purushottam W Laud and Joseph G Ibrahim. Predictive model selection. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):247–262, 1995.

[43] Jonathan W Pillow and James G Scott. Fully Bayesian inference for neural models with negative-binomial spiking. In *NIPS*, volume 1, pages 1907–1915, 2012.

[44] John Geweke. *Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA, 1991.

[45] Connecticut Data Collaborative. *Data By Topic*, 2020. URL `http://data.ctdata.org/data_by_topic#demographics`.

[46] Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. *shiny: Web Application Framework for R*, 2020. URL `https://CRAN.R-project.org/package=shiny`. R package version 1.4.0.2.

[47] Joe Cheng, Bhaskar Karambelkar, and Yihui Xie. *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*, 2019. URL `https://CRAN.R-project.org/package=leaflet`. R package version 2.0.3.

[48] Kyle Walker. *tigris: Load Census TIGER/Line Shapefiles*, 2020. URL `https://CRAN.R-project.org/package=tigris`. R package version 1.0.

[49] Edward Parker. COVID-19 tracker, 2020. URL `https://shiny.rstudio.com/gallery/covid19-tracker.html`.

[50] Edward Parker. COVID-19 interactive mapping tool, 2020. URL `https://github.com/eparker12/nCoV_tracker`.

[51] Google. COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility, 2020.

[52] Apple. Mobility Trends Reports. `https://covid19.apple.com/mobility`, 2020.

[53] Facebook. COVID-19 Interactive Map & Dashboard. https://covid-survey.dataforgood.fb.com/?region=WORLD, 2020.

[54] Descartes Labs. Changes in Mobility. https://www.descarteslabs.com/mobility/, 2020.

[55] Michael S Warren and Samuel W Skillman. Mobility changes in response to COVID-19. *arXiv preprint arXiv:2003.14228*, March 2020. URL `https://arxiv.org/abs/2003.14228`.

[56] Cuebiq. Mobility Insights. `https://www.cuebiq.com/visitation-insights-covid19/`, 2020.