INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Zachreson C, Mitchell L, Lydeamore MJ, Rebuli N, Tomko M, Geard N. 2021 Risk mapping for COVID-19 outbreaks in Australia using mobility data. *J. R. Soc. Interface* **18**: 20200657.
https://doi.org/10.1098/rsif.2020.0657

Received: 14 August 2020
Accepted: 7 December 2020

Subject Category:
Life Sciences–Physics interface

Subject Areas:
biocomplexity, computational biology

Keywords:
mobility, infectious diseases, COVID-19, transmission risk

Author for correspondence:
Cameron Zachreson
e-mail: cameron.zachreson@unimelb.edu.au

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5250901.

THE ROYAL SOCIETY PUBLISHING

Risk mapping for COVID-19 outbreaks in Australia using mobility data

Cameron Zachreson[1], Lewis Mitchell[2], Michael J. Lydeamore[3,4], Nicolas Rebuli[5], Martin Tomko[6] and Nicholas Geard[7]

[1]School of Computing and Information Systems, The University of Melbourne, Melbourne, Australia
[2]School of Mathematical Sciences, The University of Adelaide, Adelaide, Australia
[3]Victorian Department of Health and Human Services, Government of Victoria, Melbourne, Australia
[4]Department of Infectious Diseases, The Alfred and Central Clinical School, Monash University, Melbourne, Australia
[5]School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia
[6]Melbourne School of Engineering, The University of Melbourne, Melbourne, Australia
[7]The Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Melbourne, Australia

CZ, 0000-0002-0578-4049; LM, 0000-0001-8191-1997; NR, 0000-0003-2339-974X; MT, 0000-0002-5736-4679

COVID-19 is highly transmissible and containing outbreaks requires a rapid and effective response. Because infection may be spread by people who are pre-symptomatic or asymptomatic, substantial undetected transmission is likely to occur before clinical cases are diagnosed. Thus, when outbreaks occur there is a need to anticipate which populations and locations are at heightened risk of exposure. In this work, we evaluate the utility of aggregate human mobility data for estimating the geographical distribution of transmission risk. We present a simple procedure for producing spatial transmission risk assessments from near-real-time population mobility data. We validate our estimates against three well-documented COVID-19 outbreaks in Australia. Two of these were well-defined transmission clusters and one was a community transmission scenario. Our results indicate that mobility data can be a good predictor of geographical patterns of exposure risk from transmission centres, particularly in outbreaks involving workplaces or other environments associated with habitual travel patterns. For community transmission scenarios, our results demonstrate that mobility data add the most value to risk predictions when case counts are low and spatially clustered. Our method could assist health systems in the allocation of testing resources, and potentially guide the implementation of geographically targeted restrictions on movement and social interaction.

1. Introduction

Similar to other respiratory pathogens such as influenza, the transmission of SARS-CoV-2 occurs when infected and susceptible individuals are co-located and have physical contact, or exchange bioaerosols or droplets [1,2]. Behavioural modification in response to symptom onset (i.e. self-isolation) can act as a spontaneous negative feedback on transmission potential by reducing the rate of such contacts, making epidemics much easier to control and monitor. However, COVID-19 (the disease caused by SARS-CoV-2 virus) has been associated with relatively long periods of pre-symptomatic viral shedding (approx. 5–10 days), during which time case ascertainment and behavioural modification are unlikely [3,4]. In addition, many cases are characterized by mild symptoms, despite long periods of viral shedding [5]. Transmission studies have demonstrated that asymptomatic and pre-symptomatic transmission hampers control of SARS-CoV-2 [6–8]. Pre-symptomatic and asymptomatic transmission has also been documented systematically in several residential care facilities in which

surveillance was essentially complete [9,10]. Currently, there are no prophylactic pharmaceutical interventions that are effective against SARS-CoV-2 transmission. Therefore, interventions based on social distancing and infection control practices have constituted the operative framework, applied in innumerable variations around the world, to combat the COVID-19 pandemic.

Social distancing policies directly target human mobility. Therefore, it is logical to suggest that data describing aggregate travel patterns would be useful in quantifying the complex effects of policy announcements and decisions [11]. The ubiquity of mobile phones and public availability of aggregated near-real-time movement patterns has led to several such studies in the context of the ongoing COVID-19 pandemic [12–14]. One source of mobility data is the social media platform Facebook, which offers users a mobile app that includes location services at the user's discretion. These services document the GPS locations of users, which are aggregated as origin–destination (OD) matrices and released for research purposes through the Facebook Data For Good Program. The raw data are stored on a temporary basis and aggregated in such a way as to protect the privacy of individual users [15]. Several studies have used subsets of this data for analysis of the effects of COVID-19 social distancing restrictions [16–19]. In addition, other sources of mobility data have been used to quantify the positive association between human travel, case importation, and local prevalence of COVID-19 after the disease emerged in the Chinese city of Wuhan and subsequently spread to other regions [20–22].

In this work, we complement these studies by addressing the question: to what degree can real-time mobility patterns estimated from aggregate mobile phone data inform short-term predictions of COVID-19 transmission risk? Here, we examine outbreaks and population flows approximately three orders of magnitude smaller than those investigated previously in studies focused on the Chinese context [20,21]. On this scale, it is less clear whether strong associations between mobility and transmission risk will still be observable from the available data.

To address this question, we develop a straightforward procedure to generate a relative estimate of the spatial distribution of future transmission risk based on current case data or locations of known transmission centres. To critically evaluate the performance of our procedure, we retrospectively generate risk estimates based on data from three outbreaks that occurred in Australia when there was little background transmission. We do not attempt to compute precise forecasts or predictions of case incidence, which would require a transmission model. Instead, we focus on differences in observed case counts between regions and investigate the degree to which they correlate with differences in observed mobility patterns. Our intention is to examine the utility of aggregate mobility data in generating spatial assessments of outbreak risk without precise definitions or models of disease dynamics. While we acknowledge that first-order factors determining transmission between infected and susceptible hosts may dominate local disease dynamics, the hypothesis motivating this work is that mobility between locations is an important determinant with respect to transmission over large, spatially distributed populations.

The initial wave of infections in Australia began in early March 2020, and peaked on 28 March with 469 new cases. The epidemic was suppressed through widespread social distancing measures which escalated from bans on gatherings of more than 500 people (imposed on 16 March) to a nationwide 'lockdown' which began on 29 March and imposed a ban on gatherings of more than three people. By late April, daily incidence numbers had dropped to fewer than 10 per day [23]. The outbreaks we examine occurred during the subsequent period over which these general suppression measures were progressively relaxed. One of these occurred in a workplace over several weeks, one began during a gathering at a social venue, and one was a community transmission scenario with no single identified outbreak centre, which marked the beginning of Australia's 'second wave' (which is ongoing as of August 2020). The term 'community transmission' refers to situations in which multiple transmission chains have been detected with no known links identified from contact tracing and no specific transmission centres are clearly identifiable.

In each case, we use the Facebook mobility data that were available during the early stages of the outbreak to estimate future spatial patterns of relative transmission risk. We then examine the degree to which these estimates correlate with the subsequently observed case data in those regions. Our results indicate that the accuracy of our estimates varies with outbreak context, with higher correlation for the outbreak centred on a workplace, and lower correlation for the outbreak centred on a social gathering. In the community transmission scenario without a well-defined transmission locus, we compare the risk prediction based on mobility data to a null prediction based only on active case numbers. Our results indicate that mobility is more informative during the initial phases of the outbreak, when detected cases are spatially localized and many areas have no available case data.

2. Methods

Our general method is to use an OD matrix based on Facebook mobility data to estimate the diffusion of transmission risk based on one or more identified outbreak sources. The data provided by Facebook comprise the number of individuals moving between locations occupied in subsequent 8 h intervals. For an individual user, the location occupied is defined as the most frequently visited location during the 8 h interval. More details on the raw data, the aggregation and pre-processing performed by Facebook before release, and our pre-processing steps can be found in the electronic supplementary material.

COVID-19 case data are made publicly available by most Australian state health authorities on the scale of Local Government Areas (LGAs). In these urban and suburban regions, LGA population densities typically vary from approximately 0.2×10^3 to 5×10^3 residents per km^2, but can be low as 20 residents per km^2 in the suburban fringe where LGAs contain substantial parkland and agricultural zones. The output of our method is a relative risk estimate for each LGA based on their potential for local transmission. The general method is as follows:

1. Construct the *prevalence vector* $\mathbf{p}$, a column vector with one element for each location with a value corresponding to the transmission centre status of that location. For point-outbreaks in areas with no background transmission, we use a vector with a value of 1 for the location containing the transmission centre and 0 for all other locations. For outbreaks with transmission in multiple locations, we construct $\mathbf{p}$ using the number of active cases as reported by the relevant public health agency.
2. Construct an OD matrix $\mathbf{M}$, where the value of a component M_{ij} gives the number of travellers starting their journey at

location i (row index) and ending their journey at location j (column index). To approximately match the pre-symptomatic period of COVID-19, we average the OD matrix over the mobility data provided by Facebook during the 7 day period preceding the identification of the targeted transmission centre. By averaging over an appropriate time interval, the OD matrix is built to represent mobility during the initial stages of the outbreak, when undocumented transmission may have been occurring. The choice of appropriate time interval varied by scenario, as described below.

3. Multiply the OD matrix by the prevalence vector to produce an unscaled risk vector $\mathbf{r}$ with a value for each location corresponding to the aggregate strength of its outgoing connections to transmission centres, weighted by the prevalence in each transmission centre. This is re-scaled to give the relative transmission risk for each region R_i. In other words, we treat the OD matrix as analogous to the stochastic transition matrix in a discrete-time Markov chain, and compute the unscaled vector of risk values $\mathbf{r}$ as

$$\mathbf{r} = \mathbf{M}\mathbf{p}, \tag{2.1}$$

so that $\mathbf{r}$ is approximately proportional to the average interaction rate between susceptible individuals from location i and infected individuals located in the outbreak centres. These approximate interaction rates are then re-scaled to give relative risk values R_i between 0 and 1:

$$R_i = \frac{r_i}{\sum_j r_j}. \tag{2.2}$$

For point-outbreaks, this is simply

$$R_i = \frac{M_{ik}}{\sum_j M_{jk}}, \tag{2.3}$$

where k is the column index of the single outbreak location. The numerator is the number of individuals travelling from region i to the outbreak centre, and the denominator is the total number of travellers into the outbreak centre over all origin locations j.

In addition to the typical assumptions about equilibrium mixing (in the absence of more detailed interaction data), this interpretation is subject to the assumption that the strength of transmission in each centre is proportional to the number of active cases in that location. This assumption is consistent with the observation that the majority of individuals start and end their journeys in the same locations, but there is not sufficient data to unequivocally determine the relationship between transmission risk within an area and active case numbers in the resident population of that area. Therefore, it is appropriate to think of our method as a heuristic approach to estimating transmission risk based only on qualitative information about epidemiological factors and informed by near-real-time estimates of mobility patterns. These are derived from a biased sample of the population (a subset of Facebook users), and aggregated to represent movement between regions containing of the order of 10^3 to 10^5 individuals.

2.1. Context-specific factors

Outbreaks occur in different contexts, some of which may suggest use of external data sources to infer at-risk sub-populations. Such inference can be used to refine spatial risk prediction.

For example, the workplace outbreak we investigated occurred in a meat processing facility, where the virus spread among workers at the plant and their contacts. To adapt the general method to this context, we averaged OD matrices over the subset of our data capturing the transition between nighttime and daytime locations, as an estimate of work-related travel. In addition, we examined the effect of including industry of employment statistics as an additional risk factor. In this case, we used data collected by the Australian Bureau of Statistics (ABS) to estimate the proportion of meat workers by residence in each LGA, and weighted the outgoing traveller numbers by the proportion associated with the place of origin.

The resulting relative risk value R_i is a crude estimate of the probability that an individual:

— travelled from origin location i into the region containing the outbreak centre;
— travelled during the period when many cases were pre-symptomatic and no targeted intervention measures had been applied;
— made their trip(s) during the time of day associated with travel to work; and
— were part of the specific subgroup associated with the outbreak centre (in this case, those employed in meat-processing occupations).

The variation described above is specific for workplace outbreaks in which employees are infected, but could be generally applied to any context where a defined subgroup of the population is more likely to be associated (e.g. school children, aged-care workers etc.), or in which habitual travel patterns associated with particular times of day are applicable. In principle, this approach could be used to incorporate the effects of localized intervention policies or risk factors not directly related to mobility, such as limitations on gathering size, demographic factors affecting transmission risk (i.e. age distribution), or vaccination status of subpopulations. Here, in the context of the Cedar Meatworks outbreak, we focus on an occupation-related risk factor because of its assumed relationship with mobility between home and work. In the other two scenarios we investigate, no context-specific factors are incorporated.

While we make no explicit assumptions about spatial heterogeneity of disease dynamics, our choice to integrate mobility data for the 7 days preceding each risk estimate, along with our decision to validate these estimates against raw case reports implicitly assumes spatially homogeneous temporal lags between the events associated with transmission and those corresponding to subsequent case ascertainment. To test the potential sensitivity of our results to this implicit assumption, we examined the temporal autocorrelation of the mobility matrices used in our study (electronic supplementary material, figure S3). This analysis revealed a very high level of temporal consistency in relative mobility volumes between OD pairs. Because mobility patterns are consistent in time, the risk estimate at time t is not sensitive to the particular choice of integration interval as long as that interval is at least 7 days to account for weekly fluctuations in behaviour between weekend and weekday travel.

3. Results

For each of the three outbreak scenarios, we present the mobility-based estimates of the relative transmission risk distribution, and a time-varying correlation between our estimate and the case numbers ascertained through contact tracing and testing programmes. For details of these correlation computations, see the electronic supplementary material.

3.1. Cedar Meats

3.1.1. Scenario

Cedar Meats is an abattoir (slaughterhouse and meat packing facility) in Brimbank, Victoria. It is located in the western area of Melbourne. It was the locus of one of the first sizeable

outbreaks in Australia after the initial wave of infections had been suppressed through wide-spread physical distancing interventions. Meat processing facilities are particularly high-risk work environments for transmission of SARS-CoV-2, so it is perhaps unsurprising that the first large outbreak occurred in this environment [24,25]. It began at a time when community transmission in the region was otherwise undetected. As the transmission cluster grew, it was thoroughly traced and subsequently controlled. The contact-tracing effort included (but was not limited to) intensive testing of staff, each of which required a negative test before returning to work, 14-day isolation periods for all exposed individuals, and daily follow-up calls with every close contact. The outbreak was officially recognized on 29 April, when four cases were confirmed in workers at the site and, according to media reports, Victoria Department of Health and Human Services (DHHS) informed the meatworks of these findings [26]. The outbreak was first mentioned in the daily COVID-19 updates from Victoria DHHS on 2 May, when the number of confirmed cases associated with the cluster had risen to eight [27].

The Cedar Meats outbreak began when it was introduced into the workplace, where it subsequently spread to a large number of staff, and members of their households. We therefore selected for the distribution of travellers that may have been travelling *to work* in the area of Cedar Meats during the period over which undetected transmission was likely. Specifically, we generated mobility risk maps based on trips *into* the Brimbank region for the nighttime → daytime OD matrix, averaged over the period between 21 April and 27 April 2020. We note that while there were only two SARS-CoV-2 positive cases associated with the cluster during this period (in two different areas), 43 cases were detected in the following week with infected individuals residing in 14 different locations.

As our estimate of transmission risk between Brimbank and other LGAs, we compute the risk value R_i as the proportion of individuals arriving in Brimbank from any other Victorian LGA i during the nighttime → daytime OD matrix. These values were computed with equation (2.3) and are shown as a directed network in figure 1*a*. Because the outbreak occurred in an abattoir, we also explored the effect of weighting mobility by a context-specific factor: the proportion of employed persons with occupations in meat processing (figure 1*b*).

3.1.2. Risk estimates and validation with case numbers

The geographical distribution of relative transmission risk due to mobility into Brimbank during the nighttime → daytime transition is presented in figure 2*a*, while the distribution generated by including both mobility and the proportion of meat workers in each LGA is shown in figure 2*b*.

To validate our estimate, we computed Spearman's correlation between this risk estimate for each region to the time-dependent case count for each region documented over the course of the outbreak (supplied by Victoria DHHS). We use Spearman's rather than Pearson's correlation because while we expect monotonic dependence between estimated relative risk and case counts, we have no reason to expect linear dependence or normally distributed errors. The outbreak case data were supplied as a time series of cumulative detected cases in each LGA for each day of the outbreak. Therefore, we present our correlation as a function of time from 29 April, when recorded case numbers began to increase dramatically (before 1 May, the number of affected LGAs was too small compute a confidence interval ($n \leq 3$)). As case numbers increase, correlation between our risk estimates and case numbers stabilizes at approximately 0.75 using mobility only (figure 3*a*), and at approx. 0.81 when including both mobility and meatworker proportions in the risk computation (figure 3*b*). Due to privacy limitations on release of case data, we do not present case numbers by LGA for the Cedar Meats outbreak.

3.2. The Crossroads Hotel

3.2.1. Scenario

The next scenario we examine began with a single spreading event that occurred during a large gathering at a social venue in western Sydney. While workplaces have frequently been the locus of COVID-19 clusters, many outbreaks have also been sparked by social gatherings [28,29]. In urban environments, such outbreaks can prove more challenging to trace, as the exposed individuals may be only transiently associated with the outbreak location.

The Crossroads Hotel was the site of the first COVID-19 outbreak to occur in New South Wales after the initial wave of infections was suppressed. The cluster was identified on 10 July 2020, during a period when new cases numbered fewer than 10 notifications per day. However, the second wave of community transmission in Victoria produced sporadic introductions in NSW, one of which led to a spreading event at the Crossroads Hotel [30]. Based on media reports, state contact-tracing data indicated that the cluster began on the evening of 3 July, during a large gathering [31].

Unlike the Cedar Meats cluster, the Crossroads Hotel scenario was not a workplace outbreak with transmission occurring in the same context for a sustained time period, but a single spreading event in a large social centre. For this reason, to estimate relevant mobility patterns we averaged trip numbers over all time-windows in our data (daytime → evening → nighttime → daytime) for the period of 27 June–4 July. It was also necessary to perform some pre-processing of the mobility data provided by Facebook in order to correlate case data provided by New South Wales Health to our mobility-based risk estimates due to substantial differences in the geographical boundaries used in the respective data sets (see electronic supplementary material, Technical Note). Aside from these minor differences, the method applied in this scenario is essentially the same as the one described above for the Cedar Meats outbreak. Risk of transmission in an area is assessed as the proportion of travellers who entered the outbreak location from that area (see equation (2.3)).

3.2.2. Results

Correlation of our risk estimate to the number of cases in each LGA as a function of time is shown in figure 4*a*. Heat maps of estimated risk and case numbers are shown in figure 4*b* and 4*c*, respectively. In this analysis, the available data did not explicitly identify the outbreak to which each case was associated; however, it did distinguish between cases associated with local transmission clusters and those associated with international importation. Because the Crossroads Hotel cluster was the only documented outbreak during this time, we attribute to it all cluster-associated cases during the period investigated. This assumption is anecdotally consistent

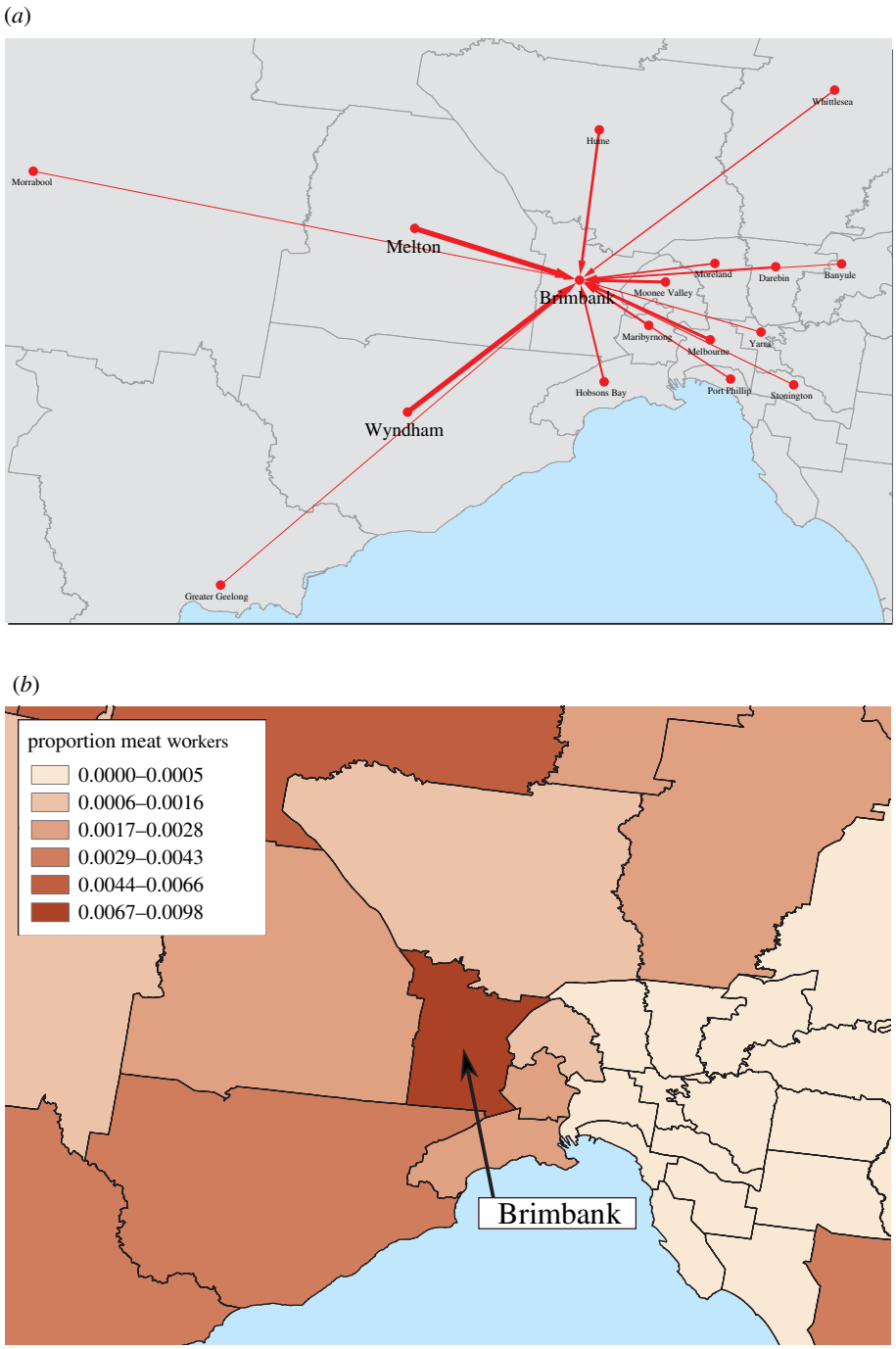

Figure 1. Risk factors for spatial proliferation of SARS-CoV-2 from the Cedar Meats outbreak in Brimbank, Victoria. (*a*) Network visualization of commuters into Brimbank during the period spanning 21 April to 27 April 2020 (arrow width is proportional to commuter numbers, with the self-loop omitted). (*b*) The proportion of employed persons in each location working in meat processing occupations as of the 2016 Census. The colour scale in (*b*) was generated using the method of Jenks natural breaks.

with media reports that specify more detailed information about the residential location of individuals associated with the outbreaks. The COVID-19 case data for New South Wales are publicly available [32].

3.3. Victoria: community transmission

3.3.1. Scenario

Both of the previous case studies considered scenarios in which a localized outbreak occurred in the context of very low or undetected community transmission. For our third case study, we consider a scenario in which community transmission had been detected across multiple suburbs in metropolitan Melbourne. This began in Victoria during June and July, after lifting of the social distancing restrictions that suppressed the initial outbreak in March. When community transmission was ascertained there were already active cases in many regions around metropolitan Melbourne, and no specific transmission centre was clearly identifiable (although transmission is thought to have originated in hotels used to quarantine arriving travellers).

On 1 June 2020, Victoria DHHS reported 71 active COVID-19 cases throughout the state, with four new cases. By 21 June, this number had increased to 121 active cases (19 new) at which point the state government initiated increased testing for community transmission and re-introduction of limits on large gatherings. Subsequently, the number of active cases increased to 645 by 6 July, and localized lockdowns were

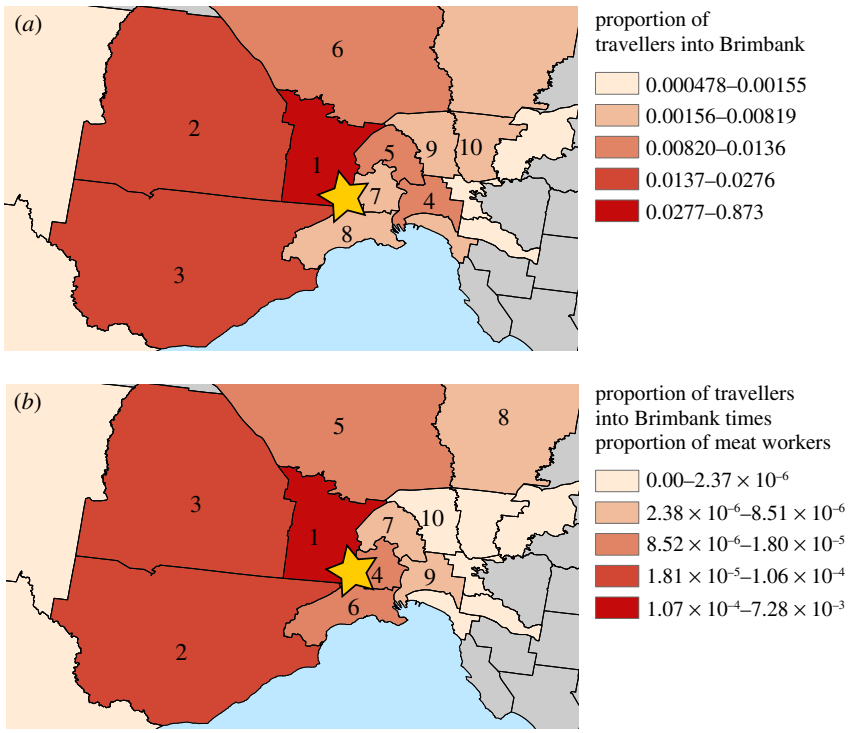

Figure 2. Regional distribution of transmission risk from the Cedar Meats outbreak in Brimbank based on (*a*) mobility into Brimbank for daytime activities, and (*b*) the proportion of employed persons with occupations in meat processing multiplied by the proportion of travellers to Brimbank shown in (*a*). The yellow stars show the approximate location of the outbreak centre. The numbers in (*a*) and (*b*) show the rank of each region with respect to the mapped quantity. The colour scales were generated using the method of Jenks natural breaks.

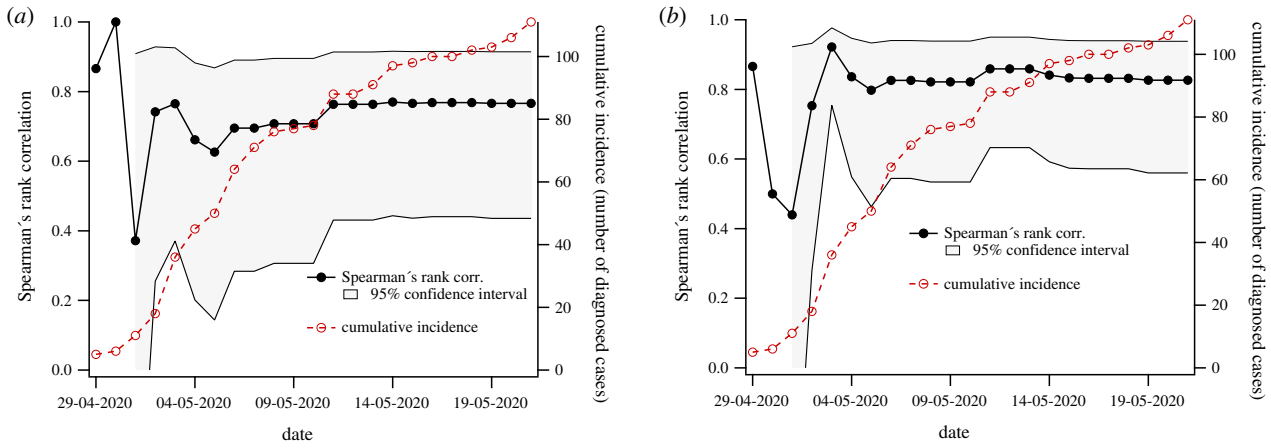

Figure 3. Correlation between risk estimates and cumulative cases as a function of time. Spearman's correlation between cases by LGA and proportion of mobility into Brimbank is shown in (*a*), while (*b*) demonstrates the effect of including employment-specific contextual factors. Black dots correspond to Spearman's correlation (left *y*-axis), and the shaded interval is the 95% CI. For reference, open red circles show total cumulative cases over all regions recorded for each day (right *y*-axis).

implemented in a set of 12 postcodes where people were asked to stay at home unless working or attending to essential activities. These targeted lockdowns were introduced in an attempt to avoid general imposition of the measures, but they were extended to the entirety of metropolitan Melbourne on 9 July, with continuing community transmission. These events are documented in the online series of daily updates provided by Victoria DHHS [33].

We examine whether the areas affected by community transmission in late June and July could have been predicted based on case numbers and mobility data that were available in early June. Our goal is to examine whether the effectiveness of mobility patterns in predicting relative transmission risk from point outbreaks can extend to community transmission scenarios in which outbreak sources are unknown.

In the community transmission scenario, as with the Crossroads Hotel outbreak, there were no clear context-dependent factors that suggested the use of other population data. In contrast to the first two scenarios, community transmission was occurring in multiple locations at the beginning of our investigation period. For each day, the unscaled risk estimate r_i is the product of the OD matrix (averaged over the preceding week) and the vector of active case numbers in each location (see equation (2.1)). Therefore, in this case, the relative risk value R_i represents the proportion of travellers into all areas containing active cases, with the

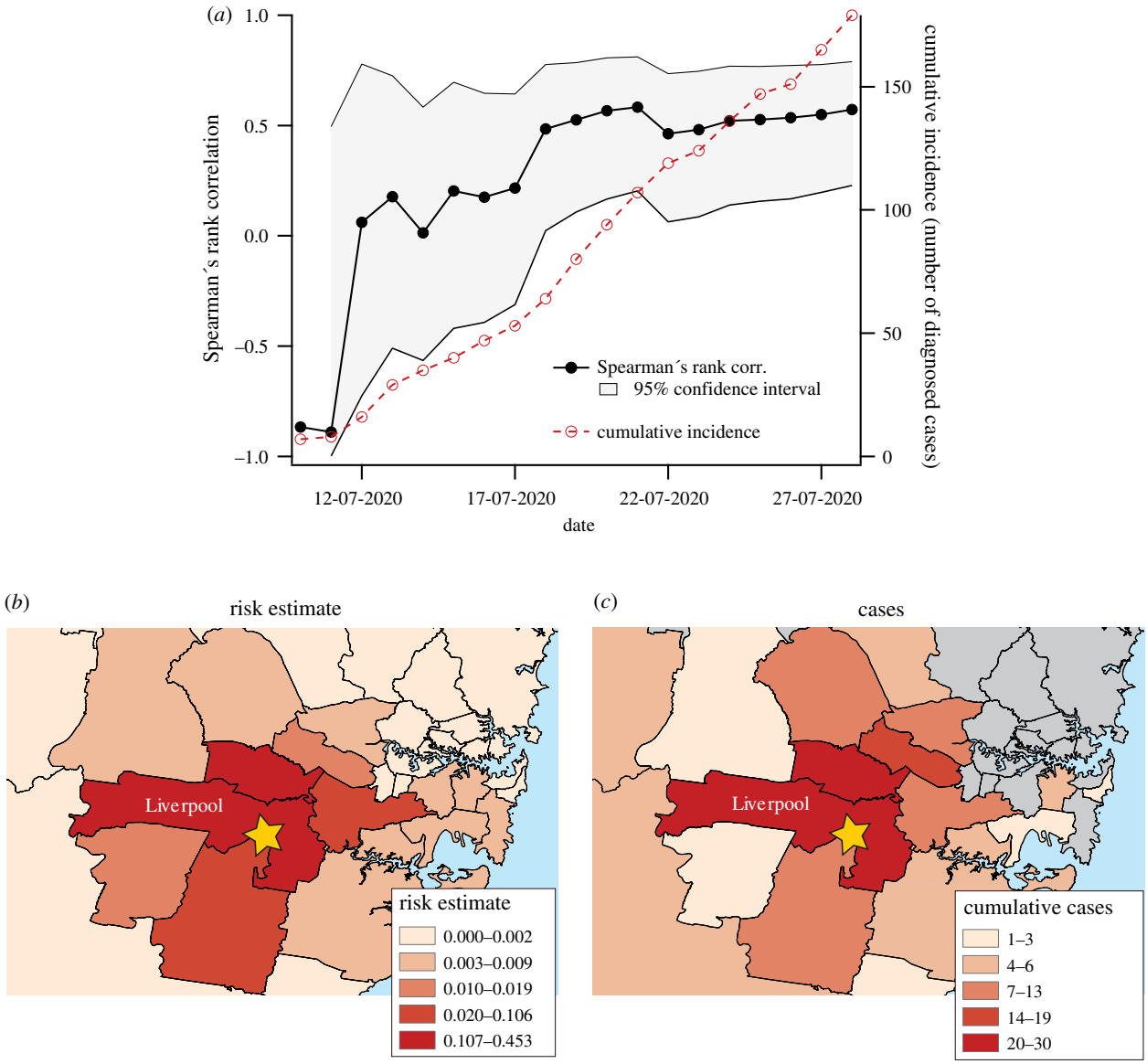

Figure 4. Comparison between estimated relative risk distribution and cluster-related case numbers in New South Wales from 12 to 28 July. Spearman's rank correlation as a function of time is shown in (*a*). The spatial distribution of estimated relative risk computed based on mobility data recorded for the week ending on 4 July is shown in (*b*), while the total number of cluster-associated cases in each LGA as of 28 July is shown in (*c*). The yellow star in (*b*) and (*c*) indicates the location of the outbreak centre, the Crossroads Hotel located in Liverpool, NSW. Colour scales in (*b*) and (*c*) were generated using the method of Jenks natural breaks.

contribution of each infected region weighted by the number of active cases (see equation (2.2)).

For this scenario, we investigate the correlation between relative risk estimates at time t, and incident case numbers (notifications) at time t', for all dates between 1 June and 21 July. We performed this more extensive analysis because it was not clear at what point in the outbreak, if any, conditions at time t would provide insight at a future time t'. In particular, we investigate if and when the incorporation of mobility data gives insight not provided by active case numbers alone.

3.3.2. Results

The results of our correlation analysis for the Victoria community transmission scenario are shown in figure 5. The correlations of incident cases (I) at time t' with active case numbers (C) and active cases combined with mobility (R) at time t are shown in figure 5*a* and 5*b*, respectively.

The added contribution of the mobility data as a function of t and t' is shown in figure 5*c*, which shows the difference between the mobility-based correlation value and the correlation based on active case numbers alone.

The values in figure 5*c* test the hypothesis that, after some delay, risk estimates incorporating mobility will correspond more closely to the future distribution of infection incidence than estimates made based only on active case numbers. Positive values support this hypothesis while values near or below zero correspond to the null hypothesis that mobility information does not improve risk estimates based on the distribution of active cases. Furthermore, by examining these values as functions of the time of risk assessment (t) and the time of case reports (t'), we gain some insight into the temporal lag between the mobility-driven diffusion of disease and the official reporting of cases. Here, the lag $(t' - t)$ between observation of risk at time t and observation of case incidence at time t' integrates all sources of delay. These include both the natural period between transmission and symptom onset, as well as the logistical delays associated

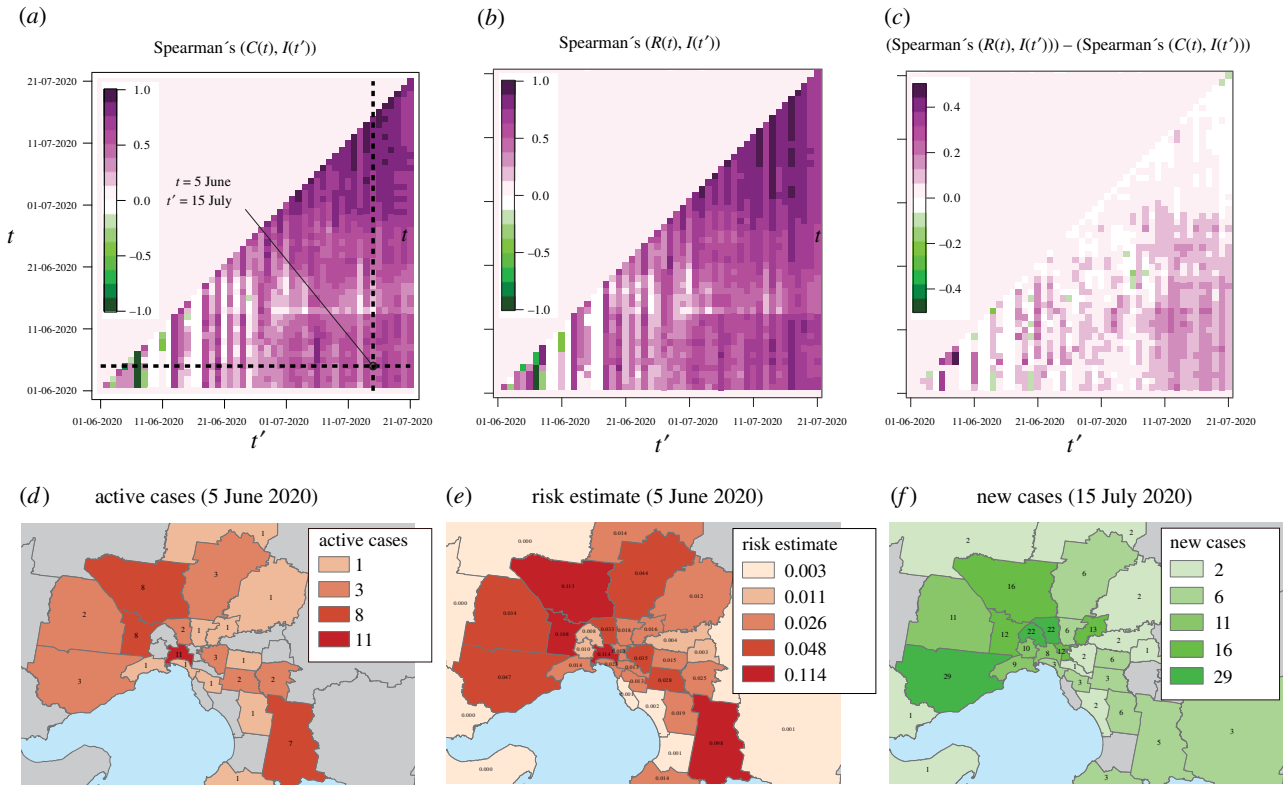

Figure 5. The contribution of mobility information to relative risk estimates in a community transmission scenario. The correlation between active cases at time t and incident cases at time t' is shown in (*a*) while the correlation of the mobility-based relative risk estimation at time t with incident cases at time t' is shown in (*b*). The benefit of including mobility information is indicated in (*c*), which shows the values plotted in (*b*) minus those plotted in (*a*). The maps in (*d–f*) correspond to the (t, t') point indicated in (*a*), and show the number of active cases on 6 June (*d*), the distribution of mobility-related relative risk on that day (*e*), and the number of incident cases on 15 July (*f*). The colour scales in (*d–f*) were generated using the method of Jenks natural breaks.

with clinical presentation, testing, ascertainment and notification. Our simple analysis does not allow us to decompose the dynamics to assess different lag components separately.

To demonstrate the geographical distribution of cases and the diffusion of risk based on mobility, figure 5*d* shows the active case counts documented for 5 June, figure 5*e* shows the corresponding distribution of transmission risk based on mobility patterns from the preceding week, and figure 5*f* shows the distribution of incident cases on 15 July. For reference, the maps in figure 5*d–f* correspond to the point indicated by the intersection of dashed lines in figure 5*a*.

4. Discussion

The goal of this study was to develop and critically analyse a simple procedure for translating aggregate mobility data into estimates of the spatial distribution of relative transmission risk from COVID-19 outbreaks. Our results indicate that aggregate mobility data can be a useful tool in estimation of COVID-19 transmission risk diffusion from locations where active cases have been identified. The utility of mobility data depends on the context of the outbreak and appears to be more helpful in scenarios involving environments where context indicates specific risk factors. The procedure we presented may also be useful during the early stages of community transmission and could help determine the extent of selective intervention measures.

In community transmission scenarios, mobility will already have played a role in determining the distribution of case counts when community transmission is detected. Our results indicate that the insight added by the incorporation of mobility data diminishes as case counts grow. However, we also observed low correlations due to stochastic effects in the Crossroads Hotel scenario. Taken together, these results indicate that there is an optimal usage window that opens when case counts are high enough for aggregate mobility patterns to shed light on transmission patterns, and closes when these transmission patterns begin to determine the distribution of active cases which then predict their own future distribution with only limited information added by considering mobility. In addition, once case counts rise sufficiently to trigger intervention policies, the local stringency of these measures and adherence to them are likely to become primary factors influencing the incidence of new cases [20,34].

Our examination of the second wave of community transmission in Victoria showed that several weeks before it was recognized, the spatial distribution of a small number of active cases was indicative of the outbreak distribution more than 30 days later when interventions were introduced. This indication improved slightly by including the diffusion of risk computed from available mobility data. Qualitatively, this observation indicates that even when case numbers were small, low-level community transmission may have already been taking place throughout the region of metropolitan Melbourne. This suggests that earlier selective lockdown measures, extending beyond the borders of regions in which cases had been identified, may have been more effective at containing transmission.

This type of relative risk estimation procedure is relevant to public health decisions relating to selective lockdown measures or the imposition of mandated infection control policies upon either the initial introduction of an infectious

disease into a susceptible population or the resurgence of a previously suppressed epidemic. Australia is currently (as of August 2020) in the early phases of the latter scenario and there is a need for policy decision frameworks aimed at preventing resurgence of the epidemic while minimizing economic consequences of further intervention. Importantly, this study focused on relatively small-scale outbreaks that all occurred within single administrative jurisdictions. For scenarios involving case importation between different administrative jurisdictions (i.e. international or interstate travel), calculation of transmission risk must take into account heterogeneity in the rates of case ascertainment within each jurisdiction, as these can vary substantially [22].

4.1. Limitations

4.1.1. Privacy, anonymity and aggregation

It is essential that the use of mobility data for disease surveillance complies with privacy and ethical considerations [11]. Due to this requirement, there will always be trade-offs between the spatio-temporal resolution of aggregated mobility data and the completeness of the data set after curation, which typically involves the addition of noise and the removal of small numbers based on a specified threshold. To help ensure users cannot be identified, Facebook removes OD pairs with fewer than 10 unique users over the 8 h aggregation period. The combination of this aggregation period with the 10-user threshold affects regional representation in the data set, particularly in more sparsely populated areas. The final product resulting from these choices contains frequently updated and temporally specific mobility patterns for densely populated urban areas, at the cost of incomplete data in sparsely populated regions. In general, increased temporal or spatial resolution will reduce trip numbers in any given set of raw data, which can have a dramatic impact on the amount of information missing from the curated numbers [35].

4.1.2. Stochastic effects

The comparison of our results from the Cedar Meats outbreak and those from the Crossroads Hotel cluster demonstrates that the utility of aggregated mobility patterns in estimation of the spatial distribution of relative risk depends on the context of the outbreak, with more value in situations involving habitual mobility such as commuting to and from work. Detailed examination of the inconsistencies between risk estimates and case data from the Crossroads Hotel outbreak indicate that small numbers of people travelling longer distances were responsible for the relative lack of correspondence in that scenario. In particular, news reports discussed instances of single individuals who had travelled from the rural suburbs to visit the Crossroads Hotel for the 3 July gathering who then infected their family members. These scenarios were not consistent with the risk predictions produced by the mobility patterns into and out of the region and exemplify the limitations of risk assessment based on aggregate behavioural data.

4.1.3. Sample bias

The mobility data provided by the Facebook Data For Good Program represent a non-uniform and essentially uncharacterized sample of the population. While it is a large sample, with aggregate counts of the order of 10% of ABS population figures, the spatial bias introduced by the condition of mobile app usage cannot be determined due to data aggregation and anonymization. While it is possible to count the number of Facebook users present in any location during the specified time intervals, it is not possible to distinguish which of those are located in their places of residence. In order to account for the (possibly many) biases affecting the sample, a detailed demographic study would be necessary that is beyond the scope of the present work. A heat map (electronic supplementary material, figure S1) of the average number of Facebook users present during the nighttime period (2:00 to 10:00) as a proportion of the estimated resident population reported by the ABS (2018 [36]) shows qualitative similarity to the spatial distributions of active cases and relative risk shown in figure 5*d*,*e*. This dual correspondence suggests the presence of common factors affecting both representation in the Facebook dataset and the risk of transmission. To investigate the potential influence of spatial sampling bias on our correlations, we performed a simple bias correction the results of which are shown in electronic supplementary material, figure S2. We did not include this bias correction as a component of our general analysis because it is unclear to what degree the correction is accurate, given a lack of detailed information on the individuals represented in Facebook user population data. That is, the bias correction we tested may have introduced different, uncharacterized biases.

4.2. Future work

On a fundamental level, mobility patterns are responsible for observed departures from continuum mechanics observed in real epidemics [37]. Over the past two decades, due to public health concern over the pandemic potential of SARS, MERS and novel influenza, spatially explicit models of disease transmission have become commonplace in simulations of realistic pandemic intervention policies [38,39]. Such models rely on descriptions of mobility patterns which are usually derived from static snapshots of mobility obtained from census data [35,40,41]. While this approach is justifiable given the known importance of mobility in disease transmission, it is also clear that the shocks to normal mobility behaviour induced by the intervention policies of the COVID-19 pandemic will not be captured by static treatments of mobility patterns. To account for the dynamic effects of intervention, several models have been developed to simulate the imposition of social distancing measures through adjustments to the strength of context-specific transmission factors [42,43]. This type of treatment implicitly affects the degree of mixing between regions without explicitly altering the topology of the mobility network on which the model is based and it is unclear whether such a treatment is adequate to capture the complex response of human population behaviour. Given the results of our analysis, the incorporation of real-time changes in mobility patterns could add policy-relevant layers of realism to such models that currently rely on static, sometimes dated, depictions of human movement.

Data accessibility. Example scripts and data used for computing risk estimates and correlations can be found in the associated GitHub repository: https://github.com/cjzachreson/COVID-19-Mobility-Risk-Mapping. However, due to release restrictions on the mobility data provided by Facebook, the OD matrices are not included as these were derived from the data provided by the Facebook Data For Good Program (random matrices are included as placeholders). The processed mobility data used in this work may be made available upon request to the authors, subject to conditions of release

consistent with the Facebook Data For Good Program access agreement. A generic implementation of the code used to re-partition OD matrices between different geospatial boundary definitions is enclosed in the electronic supplementary material, Technical Note.

Authors' contributions. All authors contributed to manuscript composition, research design and discussion. C.Z., N.G., N.R. and L.M. designed the risk estimation procedure. C.Z. and M.J.L. performed validation of risk estimates against case data. C.Z., N.R. and M.T. performed spatial data processing. C.Z. implemented data analysis scripts, and composed figures.

Competing interests. We declare we have no competing interests.

Funding. This work was supported in part by an NHMRC project grant (APP1165876), an NHMRC Centre of Research Excellence (APP1170960), and the Victoria State Government Department of Health and Human Services.

Acknowledgements. The authors acknowledge useful discussions with Nick Golding, Freya Shearer, Jodie McVernon, James McCaw and James Wood. We thank the Facebook Data For Good Program and in particular would like to acknowledge the continuous efforts of Alex Pompe in data provision and technical discussion.

References

1. Kutter JS, Spronken MI, Fraaij PL, Fouchier RA, Herfst S. 2018 Transmission routes of respiratory viruses among humans. *Curr. Opin. Virol.* **28**, 142–151. (doi:10.1016/j.coviro.2018.01.001)
2. Siegel JD, Rhinehart E, Jackson M, Chiarello L, Health Care Infection Control Practices Advisory Committee. 2007 2007 guideline for isolation precautions: preventing transmission of infectious agents in health care settings. *Am. J. Infect. Control* **35**, S65. (doi:10.1016/j.ajic.2007.10.007)
3. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. 2020 The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* **172**, 577–582. (doi:10.7326/M20-0504)
4. He X *et al.* 2020 Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675. (doi:10.1038/s41591-020-0869-5)
5. Young BE *et al.* 2020 Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *JAMA* **323**, 1488–1494. (doi:10.1001/jama.2020.3204)
6. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D, Fraser C. 2020 Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936. (doi:10.1126/science.abb6936)
7. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. 2020 Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493. (doi:10.1126/science.abb3221)
8. Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. 2020 Presymptomatic transmission of SARS-CoV-2–Singapore, January 23–March 16, 2020. *Morb. Mortal. Wkly Rep.* **69**, 411. (doi:10.15585/mmwr.mm6914e1)
9. Arons MM *et al.* 2020 Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* **382**, 2081–2090. (doi:10.1056/NEJMoa2008457)
10. Kimball A *et al.* 2020 Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility—King County, Washington, March 2020. *Morb. Mortal. Wkly Rep.* **69**, 377. (doi:10.15585/mmwr.mm6913e1)
11. Buckee CO *et al.* 2020 Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146. (doi:10.1126/science.abb8021)
12. Pepe E, Bajardi P, Gauvin L, Privitera F, Lake B, Cattuto C, Tizzoni M. 2020 COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Sci. Data* **7**, 1–7. (doi:10.1038/s41597-020-00575-2)
13. Martín-Calvo D, Aleta A, Pentland A, Moreno Y, Moro E. 2020 Effectiveness 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. In *Technical Report*.
14. Bourassa K, Sbarra D, Caspi A, Moffitt T. 2020 Social distancing as a health behavior: county-level movement in the United States during the COVID-19 pandemic is associated with conventional health behaviors.
15. Maas P, Iyer S, Gros A, Park W, McGorman L, Nayak C, Dow PA. 2019 Facebook disaster maps: aggregate insights for crisis response & recovery. In *ISCRAM*.
16. Bonaccorsi G *et al.* 2020 Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 15 530–15 535. (doi:10.1073/pnas.2007658117)
17. Lee K, Sahai H, Baylis P, Greenstone M. 2020 Job loss and behavioral change: the unprecedented effects of the India lockdown in Delhi. University of Chicago, Becker Friedman Institute for Economics Working Paper (2020-65).
18. Holtz D *et al.* 2020 Interdependence and the cost of uncoordinated responses to COVID-19.
19. Galeazzi A, Cinelli M, Bonaccorsi G, Pierri F, Schmidt AL, Scala A, Pammolli F, Quattrociocchi W. 2020 Human mobility in response to COVID-19 in France, Italy and UK. (http://arxiv.org/abs/200506341)
20. Kraemer MU *et al.* 2020 The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497. (doi:10.1126/science.abb4218)
21. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. 2020 Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 1–5. (doi:10.1038/s41586-020-2284-y)
22. Niehus R, De Salazar PM, Taylor AR, Lipsitch M. 2020 Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. *Lancet Infect. Dis.* **20**, 803–808. (doi:10.1016/S1473-3099(20)30229-2)
23. Coronavirus (COVID-19) at a glance—10 August 2020. https://www.health.gov.au/resources/publications/coronavirus-covid-19-at-a-glance-10-august-2020 (accessed 13 August 2020).
24. Dyal JW *et al.* 2020 COVID-19 among workers in meat and poultry processing facilities—19 States, April 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 557–561. (doi:10.15585/mmwr.mm6918e3)
25. Richmond CS, Sabin AP, Jobe DA, Lovrich SD, Kenny PA. 2020 Interregional SARS-CoV-2 spread from a single introduction outbreak in a meat-packing plant in northeast Iowa. *medRxiv*. (doi:10.1101/2020.06.08.20125534)
26. Sim T. 2020 First Cedar Meats COVID-19 case confirmed on 2 April. https://www.beefcentral.com/processing/first-cedar-meats-covid-19-case-confirmed-on-2-april/ (accessed 11 August 2020).
27. Coronavirus update for Victoria—02 May 2020. https://www.dhhs.vic.gov.au/coronavirus-update-victoria-02-may-2020 (accessed 11 August 2020).
28. Streeck H *et al.* 2020 Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *medrxiv*. (doi:10.1101/2020.05.04.20090076)
29. Hamner L. 2020 High SARS-CoV-2 attack rate following exposure at a choir practice—Skagit County, Washington, March 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 606–610. (doi:10.15585/mmwr.mm6919e6)
30. COVID-19 Weekly Surveillance in NSW, Epidemiological Week 31, Ending 1 August 2020. https://www.health.nsw.gov.au/Infectious/covid-19/Documents/covid-19-surveillance-report-20200801.pdf (accessed 11 August 2020).
31. Aubusson K, Visentin L. 2020 Fears of further spread as Crossroads Hotel virus cases become infectious within a day. https://www.smh.com.au/national/nsw/fears-of-further-spread-as-sydney-s-crossroads-coronavirus-cases-become-infectious-within-a-day-20200715-p55cds.html (accessed 11 August 2020).
32. NSW COVID-19 cases by location and likely source of infection; 2020. https://data.nsw.gov.au/data/dataset/nsw-covid-19-cases-by-location-and-likely-source-of-infection (accessed 11 August 2020).

33. Updates about the outbreak of the coronavirus disease (COVID-19); 2020. https://www.dhhs.vic.gov.au/coronavirus/updates (accessed 11 August 2020).
34. Sen S, Karaca-Mandic P, Georgiou A. 2020 Association of stay-at-home orders with COVID-19 hospitalizations in 4 states. *JAMA* **323**, 2522–2524. (doi:10.1001/jama.2020.9176)
35. Fair KM, Zachreson C, Prokopenko M. 2019 Creating a surrogate commuter network from Australian Bureau of Statistics census data. *Sci. Data* **6**, 1–14.
36. 1410.0 Data by Region, 2013-18; 2019. https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1410.02013-18?OpenDocument (accessed 12 August 2020).
37. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. 2006 Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451. (doi:10.1126/science.1125237)
38. Germann TC, Kadau K, Longini IM, Macken CA. 2006 Mitigation strategies for pandemic influenza in the United States. *Proc. Natl Acad. Sci. USA* **103**, 5935–5940. (doi:10.1073/pnas.0601266103)
39. Zachreson C, Fair KM, Harding N, Prokopenko M. 2020 Interfering with influenza: nonlinear coupling of reactive and static mitigation strategies. *J. R. Soc. Interface* **17**, 20190728. (doi:10.1098/rsif.2019.0728)
40. Moss R, Naghizade E, Tomko M, Geard N. 2019 What can urban mobility data reveal about the spatial distribution of infection in a single city? *BMC Public Health* **19**, 1–16. (doi:10.1186/s12889-019-6968-x)
41. Cliff OM, Harding N, Piraveenan M, Erten EY, Gambhir M, Prokopenko M. 2018 Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: an agent-based modelling approach. *Simul. Model. Pract. Theory* **87**, 412–431. (doi:10.1016/j.simpat.2018.07.005)
42. Ferguson N *et al.* 2020 Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand.
43. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. 2020 Modelling transmission and control of the COVID-19 pandemic in Australia. (http://arxiv.org/abs/200310218)