

Research



Cite this article: Stockdale JE *et al.* 2022
Quantifying transmissibility of SARS-CoV-2 and
impact of intervention within long-term
healthcare facilities. *R. Soc. Open Sci.* **9**: 211710.
<https://doi.org/10.1098/rsos.211710>

Received: 29 October 2021

Accepted: 2 December 2021

Subject Category:

Mathematics

Subject Areas:

health and disease and epidemiology/
mathematical modelling/statistics

Keywords:

COVID-19, reproduction number, outbreak
modelling, transmissibility, long-term healthcare

Author for correspondence:

Jessica E. Stockdale
e-mail: jessica_stockdale@sfu.ca

Electronic supplementary material is available
online at <https://doi.org/10.6084/m9.figshare.c.5767126>.

Quantifying transmissibility of SARS-CoV-2 and impact of intervention within long- term healthcare facilities

Jessica E. Stockdale¹, Sean C. Anderson^{1,3},
Andrew M. Edwards^{3,4}, Sarafa A. Iyaniwura^{5,7},
Nicola Mulberry¹, Michael C. Otterstatter^{6,7},
Naveed Z. Janjua^{6,7}, Daniel Coombs⁵, Caroline Colijn¹
and Michael A. Irvine^{2,7,8}

¹Department of Mathematics, and ²Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

³Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, Canada

⁴Department of Biology, University of Victoria, Victoria, Canada

⁵Department of Mathematics and Institute of Applied Mathematics, and ⁶School of Population and Public Health, University of British Columbia, Vancouver, Canada

⁷British Columbia Centre for Disease Control, Vancouver, Canada

⁸British Columbia Children's Hospital Research Institute, Vancouver, Canada

JES, 0000-0001-7984-1010; DC, 0000-0002-8038-6278;
CC, 0000-0001-6097-6708

Estimates of the basic reproduction number (R_0) for COVID-19 are particularly variable in the context of transmission within locations such as long-term healthcare (LTHC) facilities. We sought to characterize the heterogeneity of R_0 across known outbreaks within these facilities. We used a unique comprehensive dataset of all outbreaks that occurred within LTHC facilities in British Columbia, Canada as of 21 September 2020. We estimated R_0 in 18 LTHC outbreaks with a novel Bayesian hierarchical dynamic model of susceptible, exposed, infected and recovered individuals, incorporating heterogeneity of R_0 between facilities. We further compared these estimates to those obtained with standard methods that use the exponential growth rate and maximum likelihood. The total size of outbreaks varied dramatically, with range of attack rates 2%–86%. The Bayesian analysis provided an overall estimate of $R_0 = 2.51$ (90% credible interval 0.47–9.0), with individual facility estimates ranging between 0.56 and 9.17. Uncertainty in these estimates was more constrained than standard methods, particularly for smaller outbreaks informed by the population-level model. We further

estimated that intervention led to 61% (52%–69%) of all potential cases being averted within the LTHC facilities, or 75% (68%–79%) when using a model with multi-level intervention effect. Understanding of transmission risks and impact of intervention are essential in planning during the ongoing global pandemic, particularly in high-risk environments such as LTHC facilities.

1. Introduction

Early outbreaks of coronavirus disease 2019 (COVID-19) occurred in locations including public transit, places of worship, cruise ships, meat-packing plants, ski resorts, prisons and fishing vessels [1–7]. Prisons have also been highlighted as potential sources of high transmission due to housing an overcrowded ageing population with underlying health conditions [8]. Similarly, detention facilities have been noted as locations where rapid transmission could lead to hospitalizations exceeding local healthcare capacity [9]. Long-term healthcare (LTHC) facilities have also been sites of large COVID-19 outbreaks [10,11]. The virus has been observed to spread rapidly in the clustered susceptible population of LTHC facilities. Since most people in LTHC are old and frail, morbidity and mortality during an outbreak are often very high, with case fatality rates around 30% [12,13].

There has been considerable focus on estimation of the basic reproduction number R_0 in different jurisdictions during the COVID-19 pandemic [3,14–18]. Given the high fatality rate, there is a keen interest in assessing the impact of COVID-19 on LTHC facilities [13,19–22]. However, to date there has been limited work on estimation of reproductive numbers or other transmission parameters within LTHC or other high-risk facilities. Estimates of transmission parameters are useful for retrospective analysis of the efficacy of interventions—particularly when preventative measures have changed between facilities or over time within one facility—and also in planning for future outbreaks. By comparing facility-specific estimates to those from the general population, we can also understand how transmission in LTHC environments differs.

Within British Columbia (BC), Canada, there have been 99 identified and defined ‘reportable’ outbreaks, including LTHC facilities, other acute care or assisted living facilities, workplaces, correctional facilities, and religious institutions as of 21 September 2020. Fifty-three of these outbreaks were in LTHC facilities; 18 of which included more than a single case of COVID-19, were complete by 21 September and were able to be linked to capacity data. Protocols for LTHC outbreak management have been established, including implementation of a suite of interventions once an outbreak has been identified within a facility [23,24]. These interventions include isolation of active cases, isolation of all residents in the facility, increased measures around infection control, reduced barriers to testing and increased testing frequency, and increased staffing and resources. Although the precise timing of the implementation of these interventions can vary, they typically occur one to two days after identification of a resident or staff case of COVID-19.

In this work, we estimate R_0 in 18 BC LTHC COVID-19 outbreaks using a dynamic susceptible–exposed–infected–recovered (SEIR) model within a Bayesian hierarchical framework. A time-dependent infection rate in the SEIR model incorporates the implementation of COVID-19 interventions upon identification of each outbreak. Specifically, we estimate R_0 for each outbreak in the period before full outbreak interventions were put in place, as well as the strength of such interventions. The hierarchical framework allows for each LTHC outbreak to be analysed within the context of all other outbreaks in the dataset, and is therefore useful for small outbreaks where the information available from each outbreak is limited. Use of a Bayesian model also allows for the incorporation of prior knowledge into the analysis, as well as the ability to fully explore the posterior distribution of the estimates. Previous modelling work has, for example, incorporated a chain binomial transmission model within a Bayesian hierarchical model in order to incorporate multiple households into one framework [25], or sought to estimate R_0 within a Bayesian hierarchical framework using the final size of an outbreak rather than incorporating a disease dynamic model [26]. Disease dynamic models have also been used within Bayesian hierarchical frameworks for forecasting seasonal influenza [27]. In this work, we also compare the estimates from the Bayesian hierarchical model to those from several established statistical approaches for R_0 estimation, which consider each LTHC outbreak independently. We lastly investigate if the heterogeneity in our estimates of R_0 is explainable by factors such as the outbreak date or the facility capacity, or if this is likely a result of stochasticity in transmission.

2. Methods

2.1. Data

Data on reported outbreaks within BC were identified through the BC Centre for Disease Control. Those outbreaks that were identified as having taken place within a LTHC facility, that were able to be linked to facility capacity information, that were complete by the end of the study period, 21 September 2020, and that had more than one case were selected, for a total of 18 outbreaks. During the study period, no Variants of Concern had yet been identified in BC. These data consist of reported date, symptom onset date and facility location of all identified cases of COVID-19 (residents and staff). We also performed a supplementary analysis of the 30 identified single-case LTHC outbreaks meeting the above outbreak criteria, of which 25 were a staff member, 3 were residents and 2 were of an unrecorded source. These LTHC data were previously not publicly released; however, the de-identified data are now available as part of the Github repository attached to this work at github.com/semppwn/cr0eso [28].

Missing symptom onset dates (44/536 cases across all facilities) were replaced by selecting a symptom onset date uniformly at random from all known symptom onset dates within the same facility. This results in a higher chance of sampling more common symptom onset dates within each facility. Sensitivity analysis to this interpolation was performed, in which we re-sampled the missing data 100 times and explored the impact on the facility-specific R_0 estimates. The start time of each outbreak was selected as the earliest symptom onset date within all identified cases in that outbreak. The reported date for each outbreak was recorded as the earliest reported date among all reported cases in the facility, since in BC one case in a LTHC facility is sufficient to be labelled as an outbreak.

Additional covariate data concerning the LTHC facilities included in this study were obtained from the Office of the Seniors Advocate, BC, website [29]. Where available, we compiled data on various factors for each facility: average resident age, average resident stay, direct care hours, facility capacity, number of recorded disease outbreaks, number of lodged complaints, proportion of residents dependent on assistance for their daily activities, age of facility, and accreditation status. These data are from 2018/19, the most recent year available. This was combined with two additional factors concerning the outbreaks: the outbreak reported date and identity of the initial COVID-19 symptomatic case (staff, resident or unknown).

2.2. Bayesian hierarchical model

We introduce a novel modelling framework for estimating R_0 in LTHC facilities, by incorporating a dynamic compartmental SEIR model into a Bayesian hierarchical structure that includes both parameter and observation uncertainty. This framework incorporates data on each outbreak, allowing for a more robust approach to estimation than considering each outbreak separately, such that outbreaks with few cases can depend strongly on a prior distribution of R_0 . Our methodology combines the flexibility of dynamic compartmental models for transmission, allowing for estimation not just of the underlying transmission parameters but also the strength of interventions, with the benefits of Bayesian hierarchical fitting such as increased robustness of estimates and incorporation of prior knowledge.

The baseline transmissibility of COVID-19 in our model was assumed to vary between facilities due to differences in layout, contact structure, and demographics of the facilities. Within BC, standard policies seeking to reduce transmission were implemented across the province [23,30]. We therefore modelled the interventions as having equal strength in all facilities under the assumption that implementation occurred within one day of the first reported case. We also performed a sensitivity analysis in which the intervention strength was allowed to vary between facilities.

2.2.1. Transmission model

We begin with a standard SEIR model where the transmission term in a facility is dependent on the time since identification of the first case in the facility. An individual in facility k starts out as susceptible to COVID-19 (S_k), following an infection they transition to an exposed group (E_k), after an incubation period the individual transitions to an infected group (I_k), before finally transitioning to a recovered group; S_k , E_k and I_k represent the number of individuals in each group. Due to the short nature of a facility outbreak, imported and exported cases were not considered. Each outbreak was considered to be initiated by a single infectious individual. During the period of this study, general transmission in the community was low, giving a small risk of further importation over a short period. We incorporated uncertainty into the facility population sizes in part to account for staff, but otherwise did not model staff–resident contact any differently from resident–resident contact.

The model equations can be represented as

$$\begin{aligned}\frac{dS_k}{dt} &= -\frac{R_k(t)\gamma}{N_k} S_k I_k, \\ \frac{dE_k}{dt} &= \frac{R_k(t)\gamma}{N_k} S_k I_k - \sigma E_k\end{aligned}$$

and

$$\frac{dI_k}{dt} = \sigma E_k - \gamma I_k,$$

where $R_k(t)$ is the effective reproduction number in facility k at time t , γ is the recovery rate, σ is the incubation rate (the rate at which latent individuals become infectious) and N_k is the population size which is approximated by the capacity of facility k . After detection of the first case in a facility, interventions were implemented aiming to reduce the infectiousness within the outbreak cluster. We modelled this with an $R_k(t)$ term that incorporates the initial reproduction number $R_{0,k}$ in facility k , and the modified reproduction given intervention. Assuming that outbreaks are eventually brought under control leading to $R_k(t) < 1$, $R_k(t)$ can be defined using the form

$$R_k(t) = \begin{cases} R_{0,k} & t \leq \tau_k \\ R_{0,k} e^{\zeta(\tau_k - t)} & t > \tau_k, \end{cases}$$

where ζ is the intervention effect rate and τ_k is the time of the first case reported within the facility.

2.2.2. Parameter model

The parameter model incorporates the above transmission model with a multi-level R_0 term that is informed by data from all LTHC facilities in the study. More concretely, for outbreak k , an $R_{0,k}$ is drawn from the distribution:

$$R_{0,k} \sim \mathcal{N}(R_0, \sigma_R^2) [0, \infty],$$

where R_0 is the population-level basic reproduction number before outbreak detection, and the standard deviation σ_R is drawn from a standard half-normal prior [31].

The literature-based priors were selected for the population-based reproduction number R_0 , the infectious period (recovery rate), and the incubation period. As these parameters represent an estimate of the population mean, their variances were characterized by the literature-based standard error, making them highly informative. The mean rate of incubation prior (σ) had a mean of 0.2 (corresponding to an incubation period of 5 days) and standard deviation 0.025 (providing a 95% CI of approx. 4.0–6.6 days) [32]. The mean rate of recovery (γ) had a prior mean of 0.125 (corresponding to an infectious period of 8 days) with a standard deviation of 0.0125 (providing a 95% CI of approx. 6.7–10.0 days). The prior for R_0 was given a mean of 3.0 with a standard deviation of 1.0 to account for a population mean that may be higher than observed in general community transmission [33]. Other parameters, including the between-facility variance in R_0 , σ_R^2 , and timing and strength of the intervention, had more uncertainty in their prior distributions. To capture uncertainty in the facility population sizes, the population size of facility k , N_k , was given a normally distributed prior with mean equal to the known facility capacity \hat{N}_k and standard deviation 10. The prior for the initial proportion of susceptible individuals $S_k(0)$ was selected such that the mean would be 99 individuals in a population of 100 with a standard deviation of one person (so the mean number of initially infective individuals is one in a population of 100).

The full Bayesian model (without the data likelihood, which is described below) is as follows; all rates have units day^{-1} :

$$\begin{aligned}R_0 &\sim \mathcal{N}(3, 1) [0, \infty], \\ \sigma_R &\sim \mathcal{N}(0, 1) [0, \infty], \\ R_{0,k} &\sim \mathcal{N}(R_0, \sigma_R^2) [0, \infty], \\ \gamma &\sim \mathcal{N}(0.125, 0.0125^2) [0, \infty], \\ \sigma &\sim \mathcal{N}(0.2, 0.025^2) [0, \infty], \\ N_k &\sim \mathcal{N}(\hat{N}_k, 10^2) [0, \infty], \\ S_k(0) &\sim \mathcal{N}(0.99, 0.01^2) [0, \infty], \\ \tau_k &\sim \mathcal{N}(0.1, 1) [0, \infty] \\ \zeta &\sim \mathcal{N}(0.1, 0.1^2) [0, \infty].\end{aligned}$$

and

The likelihood is constructed by treating the modelled daily incidence for each facility as a Poisson-distributed random variable, with the rate given as the rate of transition from the E_k to I_k class ($u_k(t)$). Data on reported onset of symptoms on a given day for each facility ($C_k(t)$) were used as opposed to date of test, as often tests were conducted en masse at a facility once a primary case had been identified. Although this does not explicitly model presymptomatic transmission, which has been found significant for COVID-19 [34], we incorporated additional variability in our incubation and infectious period priors to account for this. And indeed, if onset of infectiousness consistently lags onset of symptoms, our outbreaks would be shifted in time but otherwise unchanged. Cases were therefore sampled according to the following to construct the likelihood:

$$C_k(t) \sim \text{Poi}(u_k(t)).$$

Sampling of the posterior was performed with no U-turn sampling (NUTS) [35] using four chains, a warm-up period of 1000 samples, followed by 1000 iterations per chain. Visual inspection of the chains, pair-plots, Gelman–Rubin (split-chain- \hat{R}) statistic, and the presence of divergent transitions were used to assess convergence and mixing [36]. All Bayesian analysis was conducted using the Stan library [37], and our model is provided as an R package accompanying this article at github.com/semppwn/cr0eso [28].

As a sensitivity analysis to understand the potential heterogeneity of the effect of intervention, the model was extended to include a hierarchical structure for the intervention effect ζ . The hyper-priors were kept the same as the ζ prior in the original model, with inter-facility variance σ_ζ drawn from a standard half-normal prior.

Counterfactual scenarios were performed on all facilities by drawing a set of parameters from the posterior and then setting the intervention effect ζ to zero for the counterfactual parameter set. We numerically solved the SEIR model for each of the 4000 parameter sets and sampled from the model to produce a difference in total incidence between the two scenarios.

Validation of the hierarchical Bayesian procedure was also performed using simulated data. We simulated 30 outbreaks from an SEIR model, each with $R_{0,k}$ drawn from a $\mathcal{N}(3, 9)$ distribution truncated at 0, Poisson-drawn incidence, and a fixed intervention implementation. Each $R_{0,k}$ was then estimated within the Bayesian hierarchical framework described above.

2.3. Comparison with established R_0 estimation methods

Established methods for estimating R_0 include calculation from the attack rate, which is the total proportion of a fixed population that contracted the disease during an outbreak [38]. Here, we estimated the attack rate as the total number of reported cases in the outbreak divided by the maximum capacity of the facility. The relationship between attack rate A and R_0 , under the model of the general deterministic epidemic, is given through the transcendental equation [39,40]

$$\log(1 - A) = -R_0 A. \quad (2.1)$$

We estimated $R_{0,k}$ for each facility from the attack rate using equation (2.1) with $A = A_k$ and $R_0 = R_{0,k}$. As a sensitivity analysis, the number of exposed individuals (the denominator in the attack rate) was varied between 85% and 115% of the maximum capacity for each facility. However, this approach does not take into account that interventions were put in place during the outbreak, and so resulting estimates are not true basic reproduction numbers but instead represent an averaging across the entire outbreak.

The basic reproduction number in each outbreak was estimated using two other approaches which consider the early growth period of an outbreak only: from the exponential growth (EG) rate [41] and from maximum likelihood (ML) [42]. These methods require an assumed generation time, for which a Gamma distribution with mean 5.2 and standard deviation 1.73, as estimated for COVID-19 in [43], was used. The EG method calculates the exponential growth rate r_k from the initial exponential period of outbreak k , which we define as the time from symptom onset of the first case to the maximum incidence day (if multiple days have maximum incidence, we pick the latest). Poisson regression is used to estimate r_k , to account for the fact that the incidence data are integer valued. $R_{0,k}$ is then calculated as $1/M(-r_k)$, where M is the moment generating function of the generation time distribution. The ML method assumes that the offspring distribution is Poisson distributed with expectation $R_{0,k}$. $R_{0,k}$ is estimated by maximizing the resulting Poisson likelihood, given daily incident case counts. This estimation is also performed over an initial exponential period only, which we define as in the exponential growth method. Both the EG and ML approaches were implemented using the R0 library in R [44].

We performed an additional analysis using the number of single-case outbreaks (which could not be incorporated into the hierarchical model, EG or ML analyses due to their small size) to obtain a single R_0 estimate for BC LTHC outbreaks in the study period, by relating R_0 to the observed number of outbreaks which did not lead to any within-facility transmission. Details of this analysis are provided in electronic supplementary material, section S1.

2.4. Correlations with outbreak characteristics

We calculated Pearson correlation coefficients (PCCs) and corresponding 95% confidence intervals between the facility-specific $R_{0,k}$ values estimated from the hierarchical model and various other covariate data concerning the LTHC facilities as listed in §2.1.

All analyses were run using R v. 4.0.0 [45,46].

3. Results

Data on 99 outbreaks within BC, Canada, as of 21 September 2020 were identified. Of these identified outbreaks, 53 were selected as having taken place within a LTHC facility. Of these, 4 were excluded as they had ambiguous names or were otherwise not able to be linked to the facility characteristics data, 30 were excluded as only having one case, and 1 was excluded as still ongoing, resulting in 18 identified LTHC outbreaks. Facilities were anonymously labelled alphabetically by increasing mean $R_{0,k}$ from the Bayesian hierarchical model. Outbreak duration and size varied greatly between locations with a duration of outbreak between 10 and 64 days (figure 1) and size of outbreak between 3 and 89 cases (figure 2; electronic supplementary material, table S2), leading to an overall attack rate between 2% and 86% (electronic supplementary material, table S2). Outbreak size variation showed no clear pattern by reported facility size (figure 2).

Bayesian hierarchical model fitting of R_0 found a large range of estimates by location, with 4 of 18 locations having median $R_{0,k}$ below the critical threshold of 1, but all of these having $R_{0,k} = 1$ within the 90% credible interval (figures 3 and 4; electronic supplementary material, table S2). The overall predictive R_0 mean estimate was 2.51 (90% CrI (credible interval) 0.47–9.0), and the range of $R_{0,k}$ estimates by location varied largely within this interval: from the lowest (location A) 0.56 (0.16–1.17) to the highest (location R) 9.17 (7.16–11.97). The incorporation of a hierarchical structure for the intervention strength (ζ) increased the overall estimate of R_0 to 4.07 (1.76–10.06) (electronic supplementary material, figures S1, S2 and table S2). The posterior predictive distributions of the LTHC outbreaks revealed tighter fits for larger outbreaks and more uncertainty for smaller ones, as would be expected (figure 4; electronic supplementary material, figure S2). Conversely, we did not find such uncertainty in the $R_{0,k}$ estimates for smaller outbreaks (figure 3). This highlights a strength of this hierarchical approach: in using data from larger outbreaks to regularize fitting in smaller ones, despite larger levels of uncertainty. In model validation, applying our approach to 30 simulated outbreaks, the hierarchical modelling procedure was able to recover the data-generating parameters (electronic supplementary material, figure S3).

Using the Bayesian hierarchical model, we also estimated the ‘critical time’ in each facility. That is, we took samples from the posterior of the length of time it took for the reproduction number to fall from $R_{0,k}$ to below 1, once each outbreak was reported and full-scale interventions were implemented. The critical time varied with $R_{0,k}$, but ranged between 0.0 days (for outbreaks with $R_{0,k}$ already below 1) and 5.54 days (figure 5; electronic supplementary material, table S2). There was considerable uncertainty in these estimates, however. Counterfactual modelling was conducted to estimate the impact of intervention on the total number of cases averted compared to an unhindered outbreak. Across all outbreaks, 61% (90% CrI 52%–69%) of all potential cases were estimated to have been averted (electronic supplementary material, figure S4). This corresponds to 890 cases (90% CrI 658.95–1184.05) avoided across the 18 outbreaks. Within each facility the impact of intervention varied: the highest at location I with 84% (90% CrI 67%–93%) cases averted and the lowest at location R with 9% (90% CrI –18% to 31%) cases averted (electronic supplementary material, figure S5). Under the model with multi-level intervention ζ , counterfactual modelling suggested a larger proportion of potential cases were averted: 75% (68%–79%) across all outbreaks, corresponding to 1681 cases (90% CrI 1215–1909), and mostly a result of a higher estimated proportion in the smallest outbreaks A–H (electronic supplementary material, figures S6 and S7).

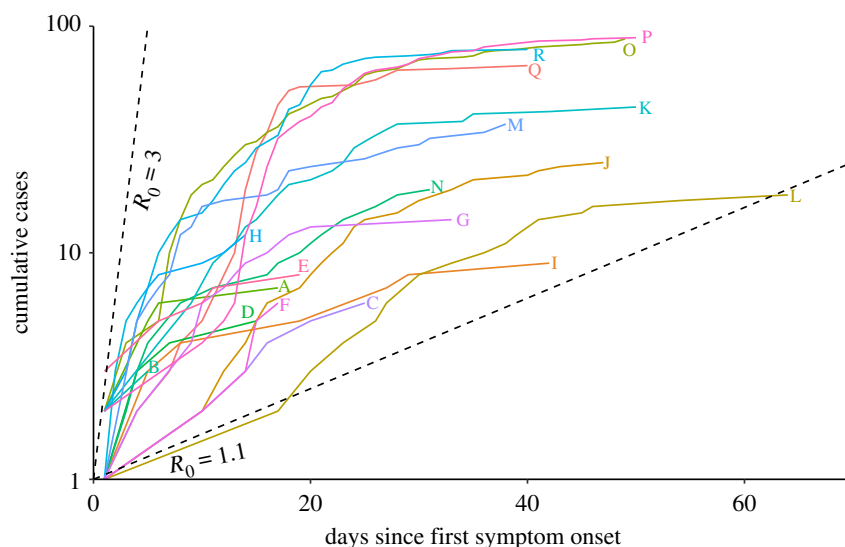


Figure 1. Cumulative cases from 18 LTHC outbreaks in BC, Canada. Cumulative number of cases for each identified outbreak is plotted against symptom onset. Early growth of cases will follow a straight line on the semi-log scale if growth is exponential. Cumulative cases are displayed as linearly increasing between successive symptom onset dates within the facility to aid in this visualization. Dashed lines represent unimpeded epidemics with R_0 values of 1.1 and 3, with a serial interval of 5 days.

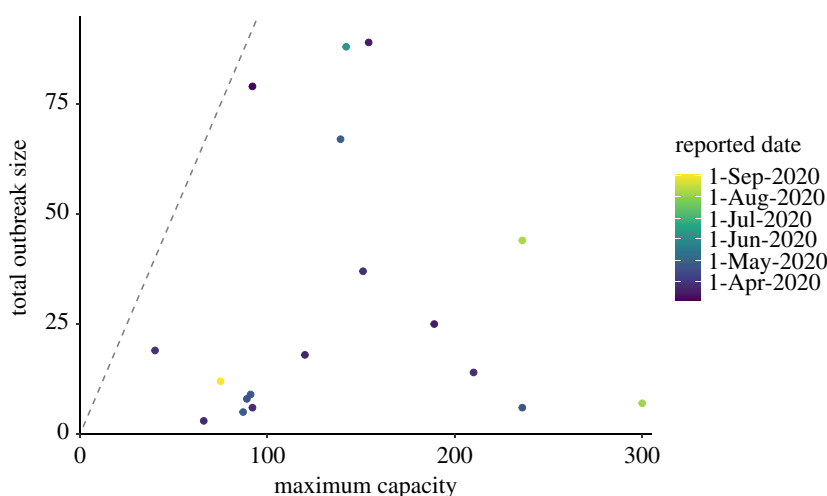


Figure 2. Total outbreak size by facility capacity of LTHC facilities in BC, Canada. Total outbreak sizes of the 18 LTHC facilities are shown against the reported maximum capacity of that facility. Dashed line is the one-to-one line. We find no clear relationship between outbreak size and capacity or reported date.

The $R_{0,k}$ estimates obtained from the EG and ML methods were quite different from those from the hierarchical model for many facilities (figure 6; electronic supplementary material, table S2), highlighting issues with fitting these (often small) outbreaks independently. Several of the smaller outbreaks (facilities A, B, D, F, G) had a large degree of uncertainty around the EG and ML estimates, which was not seen in the regularized estimates from the hierarchical model. For two outbreaks (facilities D and L) the ML method was unable to converge on an estimate: potentially due to the very fast (facility D) or slow (facility L) take-off of cases between the start of the outbreak and the maximum incidence day which we defined as the end of the exponential growth period used in fitting. The hierarchical models did not suffer from such problems. There was limited concurrence between different methods as to the outbreaks with subcritical (less than 1) $R_{0,k}$ (EG: B, E, H, I; ML: E; Bayesian hierarchical model: A, B, C, D). We see a general trend that for those outbreaks in which the hierarchical model estimated a smaller $R_{0,k}$ the EG and ML methods, unguided by any prior or hierarchical information, have large levels of uncertainty (figure 6, ordered by increasing hierarchical model $R_{0,k}$). For locations with larger $R_{0,k}$ from the hierarchical model, we see consistently lower estimates from the EG and ML methods. This may be a consequence of the

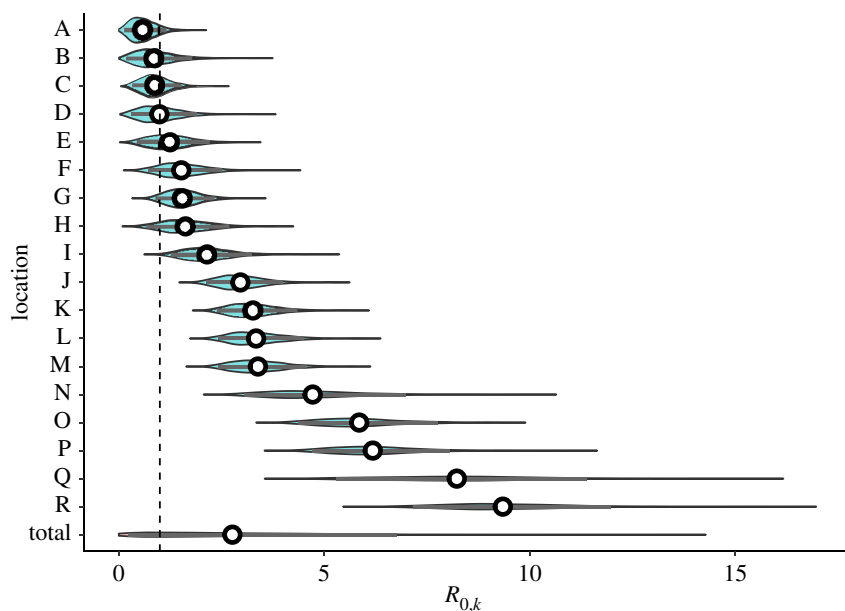


Figure 3. Hierarchical model fitting of $R_{0,k}$. Each outbreak in an LTHC facility where the outbreak was greater than one individual was incorporated into the Bayesian hierarchical SEIR model that accounted for increased control measures once the first case had been identified. The posterior distributions of the $R_{0,k}$ estimates from 4000 samples are shown, with the posterior for R_0 (total) shown below. Median values are shown as white points, with the 90% credible interval as grey bars, and the general distributions shown as density violin plots. Vertical dashed line is $R_{0,k} = 1$.

EG and ML methods working with the initial exponential period of each outbreak rather than incorporating a specific mechanism to model that interventions were put in place, unlike the hierarchical models. The impact of this may be more pronounced in those longer and larger outbreaks with higher R_0 .

To explore the impact of interpolating the missing symptom onset dates on the EG and ML $R_{0,k}$ estimates, we used a Monte Carlo procedure for repeated random sampling of the missing dates. The resulting Monte Carlo means and confidence intervals can be found in electronic supplementary material, table S1. We find that, despite the wide confidence intervals for some outbreaks in electronic supplementary material, table S2, the results are fairly robust to different interpolations of the missing symptom onset times (and many outbreaks have no missing symptom onset times). This would suggest that the majority of the uncertainty in the $R_{0,k}$ estimates does not stem from the missing data, but from some other limitation(s) of the methods.

We calculated an additional population-level estimate of R_0 from the number of outbreaks which had only one case, i.e. those in which no within-facility transmission was observed. Full results are provided in the electronic supplementary material, section S1. We found that if all 30 identified single-case outbreaks had a fully infectious individual present in the facility, the corresponding estimate of R_0 is 0.6 (95% CI 0.33–1.08). This is at odds with the large number of cases observed in many of the facility outbreaks, and suggests (a) large heterogeneity in transmission of SARS-CoV-2, as has been identified in other studies [47,48] and (b) many of the identified single cases may not have been fully infectious inside the facility. In fact, we note that 25 of the single-case outbreaks were in staff members (3 were in residents, and 2 were not recorded). It is not known whether these staff members attended the facility during their infectious period, or if an outbreak was declared out of an abundance of caution. $R_0 = 0.6$ (95% CI 0.33–1.08) therefore represents a minimum estimate from these single-case data, with the maximum estimate being $R_0 = 6.0$ (95% CI 2.0–22.8) corresponding to only the three resident cases being infectious within the facility. There is clearly considerable uncertainty here, which cannot be refined without additional information on the single-case outbreaks.

We include estimates of $R_{0,k}$ directly from the attack rate in electronic supplementary material, table S2, though it is important to note that these do not truly represent ' R_0 ' since they do not take into account that interventions were implemented and consequently represent an averaging across the whole outbreak. Nonetheless, in the absence of temporal data this may be the only way to estimate R_0 and so we include these estimates for reference. Although we would therefore not expect R_0 estimates from the attack rate to be directly comparable to those from the other methods, we did find

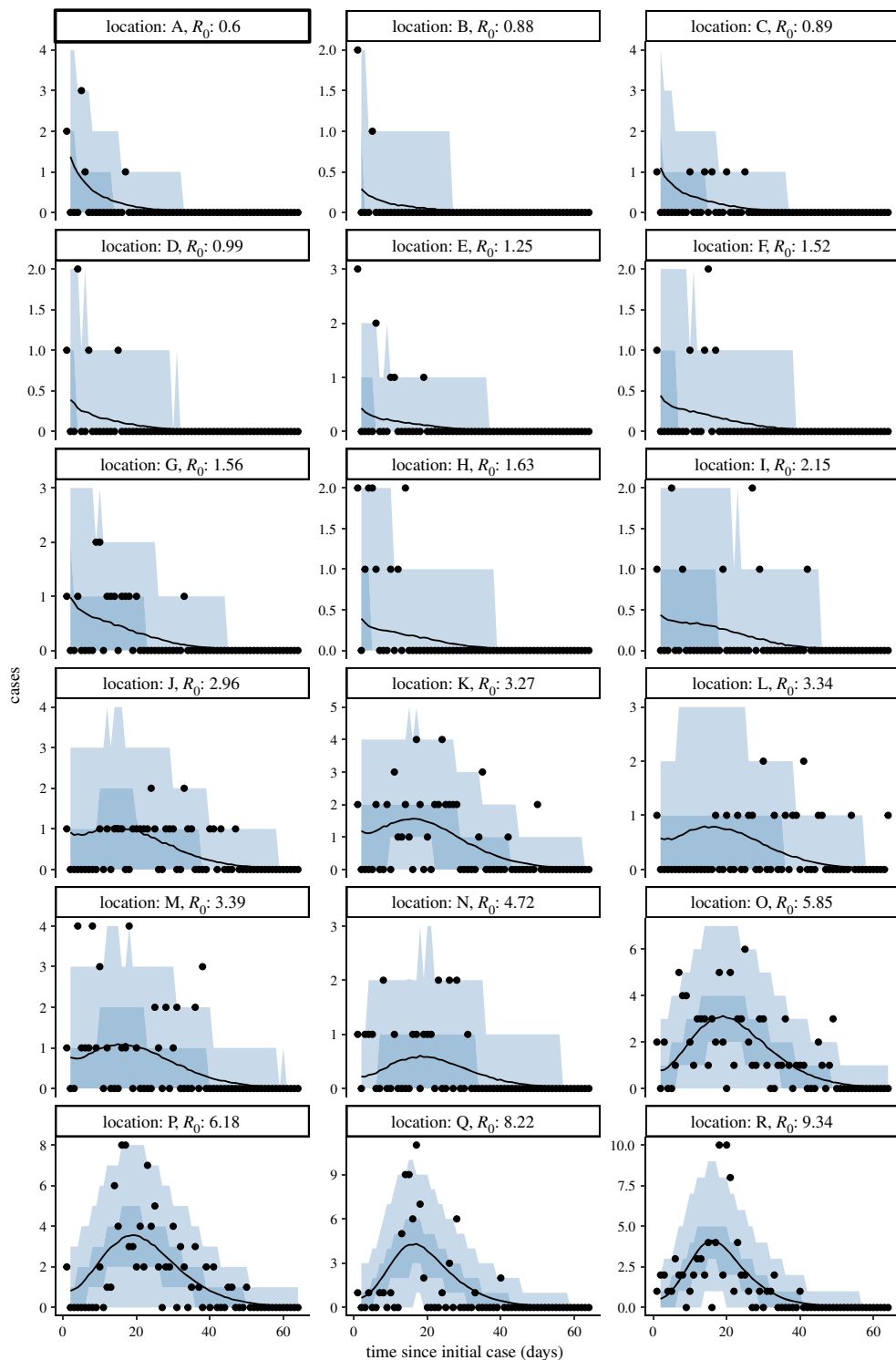


Figure 4. Bayesian hierarchical model fits to incident cases. For each LTHC outbreak of size greater than one, the posterior predictive distribution from 4000 samples is shown. Median posterior values are shown as lines with the 50% and 90% credible intervals shown as darker and lighter shaded regions, respectively. Observed cases are overlaid as black points. Facilities are ordered by increasing mean $R_{0,k}$.

significant correlation (PCC 0.94) between the $R_{0,k}$ estimates from the Bayesian hierarchical modelling and the inferred attack rates (electronic supplementary material, figure S8), which is not found between the EG and ML $R_{0,k}$ estimates and the attack rate (PCC 0.04 and 0.27). This is a result in part of the varying and high-uncertainty estimates of R_0 for smaller outbreaks under the EG and ML approaches.

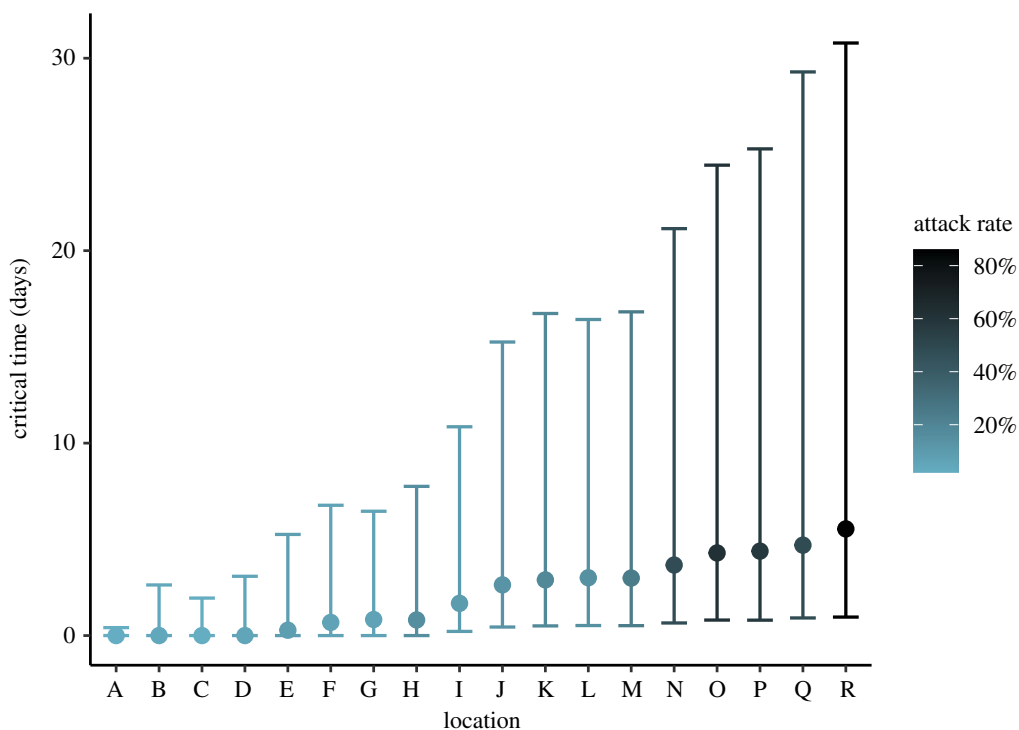


Figure 5. Posterior median estimates of the critical time from outbreak reported date to the reproduction number falling below 1. Calculated from 4000 posterior draws of $R_{0,k}$ and ζ , with error bars showing 90% credible intervals.

We found only limited correlation between our estimates of $R_{0,k}$ from the Bayesian hierarchical model and other factors regarding the LTHC facilities and their COVID-19 outbreaks, obtained from the Office of the Seniors Advocate, BC [29] (electronic supplementary material, tables S3 and S4, figure S9). No significant correlation was found between $R_{0,k}$ and the initial case in a facility (by symptom onset date) being a staff member (PCC 0.15). Though, as we do not model healthcare workers any differently from LTHC residents, our model framework is not ideal for this comparison. For the continuous variables tested, all 95% confidence intervals on the correlation coefficient included zero, with the exception of the number of lodged complaints during year 2018/19 for which we found a statistically significant negative correlation with $R_{0,k}$ (PCC -0.54 , p -value 0.04). Other small to moderate correlations were found between $R_{0,k}$ and the average stay of residents in the LTHC facility (PCC 0.38), the number of direct care hours allocated per resident per day (PCC 0.4) and, to some extent, a negative correlation between $R_{0,k}$ and the date on which the COVID-19 outbreak was reported in the facility (PCC -0.25), suggesting that later outbreaks were associated with a lower $R_{0,k}$ despite the uniform intervention policy in place in all facilities throughout the study period. However, p -values for these three tests were all above 0.05, and the sample size is relatively small; furthermore, we were unable to obtain complete data on all facilities.

4. Discussion

In this paper, we characterized variation in the estimated R_0 within a series of COVID-19 outbreaks in LTHC facilities. As R_0 is a product of infectiousness of the pathogen, contact rate and type of contact, it is important to characterize R_0 in order to understand the range of potential scenarios and total cases that could occur in future outbreaks, as well as subsequent capacity needs for healthcare provision, hospitals and intensive care units. Our Bayesian hierarchical model was able to regularize model fits even for outbreaks with a small number of cases, compared to the non-hierarchical methods for R_0 estimation which were independent between facilities and consequently suffered from high levels of uncertainty. This highlights a need for methods that consider totality of evidence. Our estimates of R_0 are specific to BC LTHC, and for many reasons could be expected to differ from estimates in the wider population. Nonetheless, the overall estimate of $R_0 = 2.51$ (0.47–9.0) agrees relatively well with published studies [15], albeit with wide uncertainty induced in part by widely varying individual facility estimates. In addition to strong

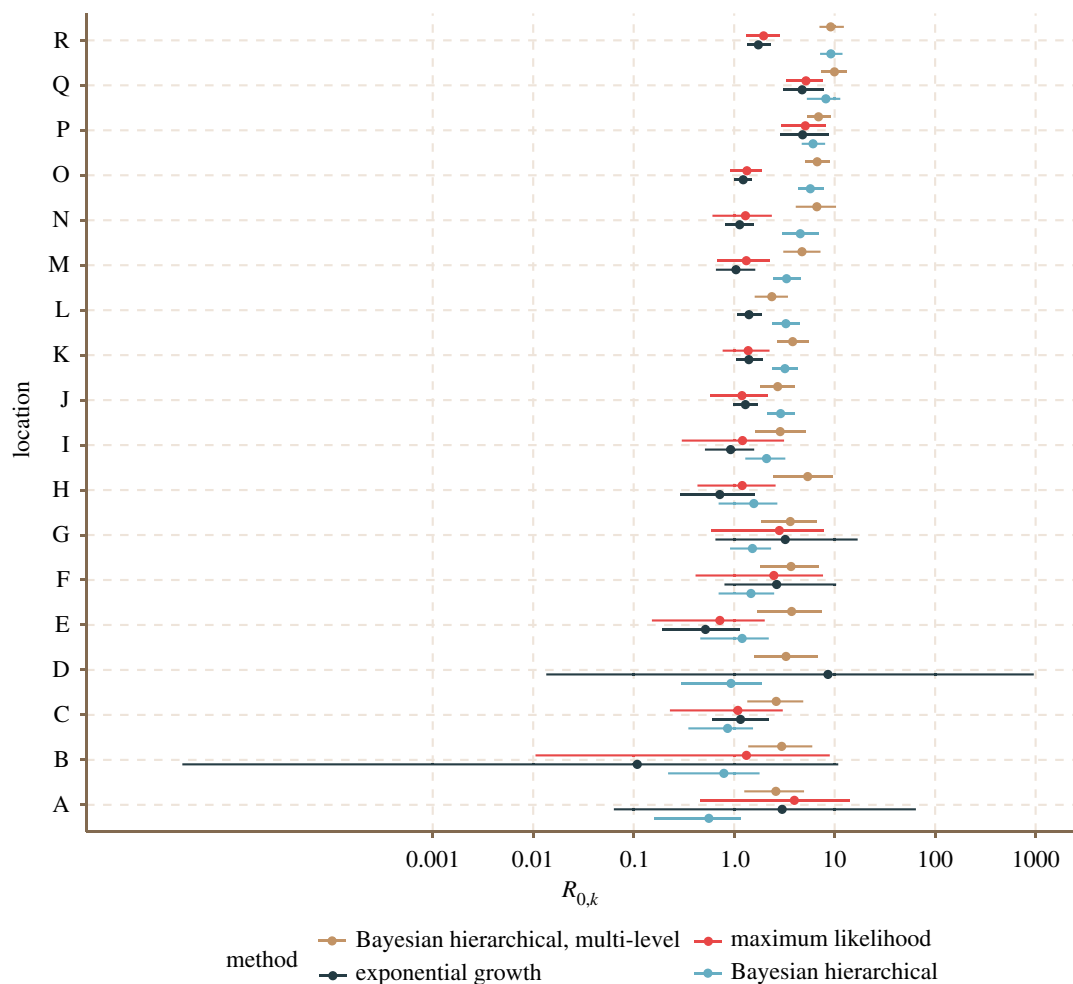


Figure 6. Comparison of $R_{0,k}$ estimates for COVID-19 outbreaks in LTHC facilities in BC, Canada, ordered from top to bottom as follows: Bayesian hierarchical multi-level median, maximum likelihood, exponential growth, Bayesian hierarchical median. Note the log-distributed axis. For each facility, $R_{0,k}$ estimates with 95% confidence interval (exponential growth and maximum likelihood) or 90% credible interval (Bayesian hierarchical models, with and without multi-level ζ term, 4000 samples) are shown.

performance, demonstrated by the model validation results, this is a flexible framework that could be adapted for different types of outbreak outside of LTHC such as schools.

We found pronounced variability in R_0 estimates by LTHC facility, despite all outbreaks occurring in facilities with the same provincial standards and with wild-type SARS-CoV-2. Variation in R_0 estimates could not be explained by facility size or whether the first infected individual was a healthcare worker. The only statistically significant relationship found was a negative correlation between R_0 and the number of complaints lodged against the facility (PCC -0.54). Further investigation would be required to explain this, but nonetheless, the majority of the variation in R_0 estimates by facility was not found to be explainable by the factors considered, nor by the missing data imputation as explored in the sensitivity analysis. Given the absence of explainable factors, this supports previous findings of the presence of super-spreader events and high natural heterogeneity between outbreaks of COVID-19 [47,48], which should be taken into account when modelling. Particularly in small outbreaks, stochastic effects can dominate, where random events can lead to a large number of secondary cases. It is important to note that our estimates do not explicitly model such events, but rather estimate the total secondary cases in aggregate. Then, the variation in estimated R_0 can be seen as explained by variation in these early events and in between-person transmissibility.

A counterfactual analysis of the scenario in which no interventions were put in place after the identification of an outbreak in each facility estimated that intervention led to 61% (52%–69%) of all potential cases being averted within LTHC facilities in BC (electronic supplementary material, figure S4). This increased to 75% (68%–79%) in the model with hierarchical intervention ζ . This suggests that fast intervention after outbreak identification is key to controlling the spread of COVID-19. Estimates

of the ‘critical time’ between implementation of interventions and the reproduction number dropping below one were all less than 6 days; however, the credible intervals were often large, up to around 30 days in the outbreaks with highest attack rate. Particularly given the high fatality rates seen in LTHC facilities, implementing strong interventions rapidly will be particularly impactful in minimizing morbidity and mortality. Although the recommended suite of interventions remained constant throughout the study period, facility outbreaks from later in the study period were weakly associated with lower critical times (PCC -0.25), suggesting some improvement in the implementation of control measures over time.

There are limitations to the study. The reported attack rate and modelled population size for LTHC facilities were based upon the maximum capacity for that facility, which may be larger than the total number of exposed individuals. Conversely, staff and family members may contribute to transmission in addition to facility residents. Although we allowed the maximum capacity size to vary to account for these factors, we lacked detailed data to inform this. Similarly, we included staff cases in our analysis but did not consider a distinction between the behaviour of staff and residents. An improved model could take this, or other factors such as presymptomatic transmission for example, into account. However, this would need to be informed by additional data. Although we estimated parameter ζ , the rate at which interventions brought $R(t)$ below 1, the manner in which intervention curtailed transmission was not explicitly explored within this study, because the exact nature of interventions and their timings were not precisely known. In our model framework, it is possible for ζ to be impacted by other factors, for example a saturation effect from a lack of available susceptible LTHC residents, rather than by intervention measures alone. However, this seems unlikely in this study, given the observed attack rates.

We considered all outbreaks as originating from a single index case in a closed population. Community prevalence remained low in BC during the period of this study so this seems a fair assumption, but it is possible that some facilities had multiple introductions. Asymptomatic cases were also not explicitly modelled in this work. It is not known if the missing symptom onset dates were caused by asymptomatic infections or data collection error, but given the low testing threshold and discretionary asymptomatic testing advised in BC LTHC facilities [49], it seems feasible that some asymptomatic cases were included in this dataset and therefore modelled the same as symptomatic cases. If many asymptomatic cases were missed however, this could be impactful on the estimates of the reproduction number. We removed all outbreaks of size one since a single data point does not contain enough information to fit the compartmental model. The additional analysis of these single-case outbreaks was not particularly informative without knowledge of how much opportunity for transmission these single cases had, exacerbated by 25/30 single case outbreaks being in staff members who may not have attended the facility while infectious. The assumption of homogeneity between outbreaks required for these calculations is also at odds with the motivations behind, and findings of, our main analysis, which suggested significant heterogeneity between locations.

The estimates of R_0 in this work (from the initial period of each outbreak) do not represent the overall level of infectivity in the outbreaks, since interventions were swiftly implemented after the initial case was discovered. The attack rate was also low in many outbreaks, leading to potentially large effects of natural stochasticity on transmission. If trying to understand the potential for future outbreaks in LTHC, it is important to consider both R_0 and ζ together. Importantly, we found that variation in R_0 was not explainable by facility size or by the initial case being a healthcare worker, but limited data meant we were also limited in what covariates could be considered. Compared to some other jurisdictions, for example many areas of the USA [10], BC has seen a relatively small number of large outbreaks. With sufficiently detailed data, this analysis could be strengthened by considering other geographical locations. Other studies have considered overdispersion of secondary cases among outbreaks, finding high levels of overdispersion among country-level outbreaks more generally [50]. It is not clear currently how much of this heterogeneity in secondary cases is due to contact patterns, age of cases, or other biological factors. As in this study, we focused only on identified outbreaks and only on LTHC facilities, population-level overdispersion caused by sampling inconsistencies and differences in individual behaviour should be minimized, compared to country-level estimates for example. A negative binomial observational model could better incorporate remaining overdispersion caused by natural individual heterogeneity, but may lead to fitting challenges compared to the more restrictive Poisson model, particularly for smaller outbreaks as explored here. Overall, the Poisson model was found to provide a reasonable fit to the data.

For future LTHC outbreaks, measures such as regular testing of staff could assist with identifying potential outbreaks as early as possible. Across the 18 LTHC outbreaks in this study, two-thirds of those with known origin had a staff member as the initial case by symptom onset. At least 83% of

those outbreaks with only 1 case had a staff member as the initial case. The analysis we have performed here could be refined with more accurate information on the total number of residents/staff per facility, as well as on differing levels of interaction between staff and residents. But also, more generally, prospective collection of data concerning factors that have been shown to impact spread of COVID-19 in other settings could help reveal sources of variation in transmissibility between different LTHC settings: including ventilation systems, staff deployment practices and patient layouts.

Despite evidence of improvement in the implementation of control measures in LTHC facilities over time, by the end of 2020 the number of reported long-term care facility outbreaks in BC increased significantly to 224 [51]. This may largely be due to increased community prevalence [51], increasing the rate of importation to LTHC facilities, but also highlights significant remaining challenges in outbreak management. Modelling an increased community prevalence of COVID-19 would require a generalization of our current approach, which does not consider repeat importations to each LTHC facility, by estimation of a rate of case importation. This is a topic for future study, but could be highly informed by the current model in which we can more reasonably assume a single point of introduction. The increase in both community prevalence and the number of facility outbreaks in BC also highlights that LTHC facilities are strongly connected to their surrounding community, and this community plays a large role in protecting LTHC facilities from outbreaks, of COVID-19 and beyond. Although our findings suggest that the suite of interventions implemented in BC LTHC facilities has been effective in significantly reducing the size of COVID-19 outbreaks, the high variability in the R_0 estimates, despite uniform controls applied and lack of correlation to other available factors, suggests that early events (before outbreak detection) and large heterogeneity in between-person transmissibility are highly impactful. With regions including BC still facing significant COVID-19 case counts despite vaccine uptake [52], combined with concerns over waning vaccine-induced immunity and weaker immune response in the elderly [53], our findings suggest that further improving controls, particularly before outbreaks begin, is key to reducing disease burden and associated mortality in LTHC facilities under increased community transmission.

Ethics. The study was approved by the Behavioural Research Ethics Board of the University of British Columbia, UBC BREB no. H20-02097. No fieldwork was conducted as part of this study. The study contains secondary use of de-identified and aggregated data collected as part of routine public health operations for quality improvement and programme evaluation, and so informed consent was not required.

Data accessibility. Data and relevant code for this research work are stored in GitHub: github.com/semprwn/cr0eso and have been archived within the Zenodo repository: <https://doi.org/10.5281/zenodo.5722322> [28]. Covariate LTHC facility data were obtained from <https://www.seniorsadvocatebc.ca/quickfacts/location> [29]. Further cumulative data on LTHC outbreaks in BC are available from <http://www.bccdc.ca/health-info/diseases-conditions/covid-19/data> [52].

Authors' contributions. J.E.S.: conceptualization, data curation, formal analysis, methodology, software, visualization, writing—original draft, writing—review and editing; S.A.: conceptualization, software, writing—review and editing; A.M.E.: conceptualization, writing—review and editing; S.I.: conceptualization, writing—review and editing; N.M.: conceptualization, writing—review and editing; M.O.: conceptualization, data curation, funding acquisition, writing—review and editing; N.J.: conceptualization, data curation, funding acquisition, writing—review and editing; D.C.: conceptualization, funding acquisition, writing—review and editing; C.C.: conceptualization, funding acquisition, writing—review and editing; M.A.I.: conceptualization, data curation, formal analysis, methodology, software, visualization, writing—original draft, writing—review and editing.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by funding from the Michael Smith Foundation for Health Research, the Canadian Institutes of Health Research (CIHR) and the Peter Wall Institute for Advanced Studies. C.C. and J.E.S. are funded by the Federal Government of Canada's Canada 150 Research Chair programme.

Acknowledgements. The authors thank Dr Rohit Vijh for advice and perspectives, and Fisheries and Oceans Canada for their support. Thank you to the British Columbia Centre for Disease Control, Fraser Health Authority and Vancouver Coastal Health Authority for sharing the data used in this analysis. Thank you also to the three anonymous reviewers and Associate Editor whose comments and suggestions helped improve and clarify the manuscript.

References

1. Shen Y *et al.* 2020 Community outbreak investigation of SARS-CoV-2 transmission among bus riders in eastern China. *JAMA Int. Med.* **180**, 1665–1671. (doi:10.1001/jamainternmed.2020.5225)
2. Hamner L *et al.* 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)
3. Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. 2020 Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis. *Int. J. Infect. Dis.* **93**, 201–204. (doi:10.1016/j.ijid.2020.02.033)

4. 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)
5. Danis K *et al.* 2020 Cluster of coronavirus disease 2019 (COVID-19) in the French Alps, 2020. *Clin. Infect. Dis.* **71**, 825–832. (doi:10.1093/cid/ciaa424)
6. Wallace M *et al.* 2020 COVID-19 in correctional and detention facilities—United States, February–April 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 587–590. (doi:10.15585/mmwr.mm6919e1)
7. Addetia A *et al.* 2020 Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with a high attack rate. *J. Clin. Microbiol.* **58**, e02107–e02120. (doi:10.1128/JCM.02107-20)
8. Kinner SA *et al.* 2020 Prisons and custodial settings are part of a comprehensive response to COVID-19. *Lancet Public Health* **5**, e188–e189. (doi:10.1016/S2468-2667(20)30058-X)
9. Irvine M *et al.* 2020 Modeling COVID-19 and its impacts on US Immigration and Customs Enforcement (ICE) detention facilities, 2020. *J. Urban Health* **97**, 439–447. (doi:10.1007/s11524-020-00441-x)
10. Van Houtven CH, Boucher NA, Dawson WD. 2020 Impact of the COVID-19 outbreak on long-term care in the United States. International Long-Term Care Policy Network, CPEC-LSE. Available from: <https://ltsccovid.org>.
11. Comas-Herrera A, Zalakain J, Litwin C, Hsu AT, Lane N, Fernández JL. 2020 Mortality associated with COVID-19 outbreaks in care homes: early international evidence. International Long-Term Care Policy Network, CPEC-LSE. Available from: <https://ltsccovid.org>.
12. McMichael TM *et al.* 2020 Epidemiology of COVID-19 in a long-term care facility in King County, Washington. *N. Engl. J. Med.* **382**, 2005–2011. (doi:10.1056/NEJMoa2005412)
13. Hsu AT, Lane N. 2020 Impact of COVID-19 on residents of Canada's long-term care homes—ongoing challenges and policy response. International Long-Term Care Policy Network, CPEC-LSE. Available from: <https://ltsccovid.org>.
14. Anderson SC *et al.* 2021 How much leeway is there to relax COVID-19 control measures?. *Epidemics* **35**, 100453. (doi:10.1016/j.epidem.2021.100453)
15. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. 2020 The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, taaa021. (doi:10.1093/jtm/taaa021)
16. Hilton J, Keeling MJ. 2020 Estimation of country-level basic reproductive ratios for novel Coronavirus (SARS-CoV-2/COVID-19) using synthetic contact matrices. *PLoS Comput. Biol.* **16**, e1008031. (doi:10.1371/journal.pcbi.1008031)
17. Tuite AR, Fisman DN, Greer AL. 2020 Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. *CMAJ* **192**, E497–E505. (doi:10.1503/cmaj.200476)
18. Donnat C, Holmes S. 2021 Modeling the heterogeneity in COVID-19's reproductive number and its impact on predictive scenarios. *J. Appl. Stat.* 1–29. (doi:10.1080/02664763.2021.1941806)
19. Logar S. 2020 Care home facilities as new COVID-19 hotspots: Lombardy Region (Italy) case study. *Arch. Gerontol. Geriatr.* **89**, 104087. (doi:10.1016/j.archger.2020.104087)
20. Stall NM, Jones A, Brown KA, Rochon PA, Costa AP. 2020 For-profit long-term care homes and the risk of COVID-19 outbreaks and resident deaths. *CMAJ* **192**, E946–E955. (doi:10.1503/cmaj.201197)
21. Burton JK *et al.* 2020 Evolution and effects of COVID-19 outbreaks in care homes: a population analysis in 189 care homes in one geographical region of the UK. *Lancet Healthy Longevity* **1**, e21–e31. (doi:10.1016/S2666-7568(20)30012-X)
22. Béland D, Marier P. 2020 COVID-19 and long-term care policy for older people in Canada. *J. Aging Soc. Policy* **32**, 358–364. (doi:10.1080/08959420.2020.1764319)
23. BC Centre for Disease Control and BC Ministry of Health. 2020 Infection prevention and control requirements for COVID-19 in long term care and seniors' assisted living. http://www.bccdc.ca/Health-Info-Site/Documents/COVID19_LongTermCare_AssistedLiving.pdf (accessed 6 July 2020).
24. Vijh R *et al.* 2021 Evaluation of a multi-sectoral intervention to mitigate the risk of SARS-CoV-2 transmission in long-term care facilities. *Infect. Control Hosp. Epidemiol.* **42**, 1–37. (doi:10.1017/ice.2020.1407)
25. Hsu C, Yen A, Chen L, Chen H. 2015 Analysis of household data on influenza epidemic with Bayesian hierarchical model. *Math. Biosci.* **261**, 13–26. (doi:10.1016/j.mbs.2014.11.006)
26. Comin A *et al.* 2010 Estimate of basic reproduction number (R_0) of low pathogenicity avian influenza outbreaks using a Bayesian approach. *Society for Veterinary Epidemiology and Preventive Medicine Proc., Nantes, France, 24–26 March 2010*, pp. 145–153.
27. Osthus D *et al.* 2019 Dynamic Bayesian influenza forecasting in the United States with hierarchical discrepancy (with discussion). *Bayesian Anal.* **14**, 261–312. (doi:10.1214/18-BA1117)
28. Irvine MA, Stockdale JE, Anderson SC. 2021 sempwn/crOeso: Published release (v0.0.0.9.002). *Zenodo*. <https://doi.org/10.5281/zenodo.5722322>.
29. Office of the Seniors Advocate, British Columbia. 2020 Long-Term Care Facilities Quick Facts Directory. <https://www.seniorsadvocatebc.ca/quickfacts/location> (accessed 30 November 2020).
30. Liu M *et al.* 2020 COVID-19 in long-term care homes in Ontario and British Columbia. *CMAJ* **192**, E1540–E1546. (doi:10.1503/cmaj.201860)
31. Gelman A, Hill J. 2006 *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
32. McAloon C *et al.* 2020 Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open* **10**, e039652. (doi:10.1136/bmjopen-2020-039652)
33. Alimohamadi Y, Taghdir M, Sepandi M. 2020 The estimate of the basic reproduction number for novel coronavirus disease (COVID-19): a systematic review and meta-analysis. *J. Prev. Med. Public Health* **53**, 151–157. (doi:10.3961/jpmph.20.076)
34. Tindale LC *et al.* 2020 Evidence for transmission of COVID-19 prior to symptom onset. *Elife* **9**, e57149. (doi:10.7554/eLife.57149)
35. Hoffman MD, Gelman A. 2014 The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623.
36. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2013 *Bayesian data analysis*. New York, NY: CRC Press.
37. Stan Development Team. 2020 RStan: the R interface to Stan. R package version 2.21.2. Available from: <http://mc-stan.org/>.
38. Lessler J, Cummings DA, Fishman S, Vora A, Burke DS. 2007 Transmissibility of swine flu at Fort Dix, 1976. *J. R. Soc. Interface* **4**, 755–762. (doi:10.1098/rsif.2007.0228)
39. Diekmann O, Heesterbeek JAP. 1989 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. New York, NY: Wiley.
40. Dietz K. 1993 The estimation of the basic reproduction number for infectious diseases. *Stat. Methods Med. Res.* **2**, 23–41. (doi:10.1177/096228029300200103)
41. Wallinga J, Lipsitch M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B.* **274**, 599–604. (doi:10.1098/rspb.2006.3754)
42. White LF *et al.* 2009 Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza Other Respir. Viruses* **3**, 267–276. (doi:10.1111/j.1750-2659.2009.00106.x)
43. Ganyani T *et al.* 2020 Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance* **25**, 2000257. (doi:10.2807/1560-7917.ES.2020.25.17.2000257)
44. Obadia T, Haneef R, Boëlle PY. 2012 The R_0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med. Inform. Decis. Mak.* **12**, 147. (doi:10.1186/1472-6947-12-147)
45. Wickham H *et al.* 2019 Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686. (doi:10.21105/joss.01686)
46. R Core Team. 2020 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>.
47. Zhang Y, Li Y, Wang L, Li M, Zhou X. 2020 Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China. *Int. J. Environ. Res. Public Health* **17**, 3705. (doi:10.3390/ijerph17103705)
48. Wang L *et al.* 2020 Inference of person-to-person transmission of COVID-19 reveals hidden

super-spreading events during the early outbreak phase. *Nat. Commun.* **11**, 5006. (doi:10.1038/s41467-020-18836-4)

49. BC Centre for Disease Control and BC Ministry of Health. 2021 COVID-19: outbreak management protocol for acute care, long-term care and seniors' assisted living settings. http://www.bccdc.ca/Health-Professionals-Site/Documents/COVID-19_Provincial_Outbreak_Management_Protocol_LTC.pdf (accessed 8 September 2021).

50. Endo A *et al.* 2020 Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67. (doi:10.12688/wellcomeopenres.15842.3)

51. BC Centre for Disease Control. 2020 British Columbia (BC) COVID-19 Situation Report Week 50: December 6–December 12, 2020. http://www.bccdc.ca/Health-Info-Site/Documents/COVID_sitrep/Week_50_BC_COVID-

19_Situation_Report.pdf (accessed 6 January 2021).

52. BC Centre for Disease Control. 2020 BC COVID-19 data. <http://www.bccdc.ca/health-info/diseases-conditions/covid-19/data#outbreak> (accessed 26 January 2021).
53. Schwarz T *et al.* 2021 Delayed antibody and T-cell response to BNT162b2 vaccination in the elderly, Germany. *Emerg. Infect. Dis.* **27**, 2174. (doi:10.3201/eid2708.211145)