

A Monte Carlo EM Analysis of COVID-19 Outbreaks in Long-Term Healthcare Facilities

William Ruth¹ and Richard Lockhart²

¹Corresponding Author - Department of Statistics and Actuarial Science , Simon Fraser University , Burnaby, BC Canada , wruth@sfu.ca

²Department of Statistics and Actuarial Science , Simon Fraser University , Burnaby, BC Canada

Abstract

We analyze a dataset of daily case counts from COVID-19 outbreaks in long-term care facilities across BC, Canada. We treat infection durations as missing data and apply the Ascent-Based Monte Carlo EM algorithm to estimate the probability of transmission and mean infection duration, as well as a third parameter governing the effectiveness of outbreak management.

We use truncated importance sampling to generate Monte Carlo samples for our analysis. We use an innovative application of a simple EM identity to compare results across multiple starts of our algorithm. We also investigate the effect of multiple imputation on our dataset.

Finally, we present the results of our analysis, including an estimate of the basic reproduction number, \mathcal{R}_0 , and uncertainty quantification for each of our parameter estimates.

I need to be consistent about what I call the estimates which are retained on each imputed dataset. I variously call them plausible estimates, estimates with large likelihood, promising estimates, retained estimates, probably others too. Right now, I like retained estimates, but that might change. Watch for this when reading to edit.

What do I need to define, and has everything been defined? A ctrl+f search suggests that I don't need to define the usual "missing data distribution", but a more thorough reading might prove otherwise.

Before finalizing, check for: **hl**, **textbf**, **textcolor**, (Citation Needed).

1 Introduction

The COVID-19 pandemic has been a major world event, and the subject of a vast amount of research. The complexity of a global pandemic is far beyond the scope of a single

study. We instead focus on a single location, the province of British Columbia (BC) in Canada, and a single population, people living in long-term healthcare (LTHC) facilities. Specifically, we study the dynamics of 53 outbreaks within LTHC facilities across the province of BC. We estimate the infectivity and case duration, as well as a parameter which measures the effectiveness of outbreak mitigation protocols within the facilities. Our model also facilitates estimation of the famous ‘basic reproduction number’, or \mathcal{R}_0 . We report estimates for each of these parameters, as well as associated measures of uncertainty.

Our work consists of a frequentist analysis of the dataset provided by Stockdale et al. (2022). We build upon their work, and also that of Vijn et al. (2022), by incorporating information from both the single-case outbreaks and the entire trajectories of outbreaks with multiple cases. A key feature of these outbreaks, the case recovery times, is not available. We therefore use a method based on the EM algorithm to maximize the likelihood of our observed data. More specifically, we use the Ascent-Based Monte Carlo EM algorithm of Caffo et al. (2005), along with the Truncated Importance Sampling method of Ionides (2008) to fit our model. We also present an innovative application of an identity from Chan and Ledolter (1995) to allow comparison of results across multiple initializations. Finally, the dataset of Stockdale et al. contains a number of missing infection times, which they handle by multiple imputation. We therefore repeat our analysis on each of the 100 imputations and investigate the relative contributions of different sources of variability.

Computation for our analysis is done in `Julia` (Bezanson et al., 2017). Our code is available in a Github repository (Citation Needed).

The rest of this paper (chapter) is organized as follows. Section 2 describes the dataset in more detail, and discusses the analyses of Vijn et al. (2022) and Stockdale et al. (2022). Section 3 gives our probabilistic model for the data, and Section 4 details how we fit this model to the data. Section 5 gives our results, and finally, Section 6 gives our conclusions, as well as some discussion of the limitations and implications of our work.

2 Related Analyses and Data

Vijn et al. (2022) analyze a dataset consisting of COVID outbreaks in long-term healthcare (LTHC) facilities across British Columbia (BC), as well as several features of the facilities and their operations. These authors perform a negative binomial GLM to measure the association between LTHC facilities’ features and the attack rates of the corresponding outbreaks. Population size is assumed to be equal to the number of beds at the facility, plus the number of staff on payroll. A limitation of this analysis is that the attack rate only accounts for the final outbreak size, not the trajectory of daily case counts.

A different analysis was carried out by Stockdale et al. (2022). Their dataset consists of daily case counts from 53 outbreaks across different facilities. This dataset has a number of missing infection times (i.e., instances where we know a case occurred but not when it started). Stockdale et al. address this missingness by multiple imputation, with missing

infection times sampled from other infection times in the same outbreak. They then repeat this imputation 100 times, producing 100 versions of the dataset. After imputing missing values, Stockdale et al. fit a hierarchical Bayesian model to estimate a local reproduction number for each LTHC facility. They also included an ‘intervention parameter’, which governs the rate at which efforts to slow the spread of an active outbreak reduce the value of the local reproduction number over the course of that outbreak. A limitation of this work is the exclusion of single-case outbreaks from their analysis. The analysis by Stockdale et al. (2022) focuses on directly modelling \mathcal{R}_0 , which does not incorporate information from outbreaks which never advance past the first case. In particular, of the 53 outbreaks in the dataset, 35 have only one case. We therefore expect the estimated reproduction number of Stockdale et al. to be biased high, as the omission of short outbreaks will inflate the expected rate of disease spread.

Note that there is some uncertainty around the nature of the single-case outbreaks. Most of these cases were among facility staff (see Supplemental Material to Stockdale et al., 2022). It is possible that staff may have self-isolated at home upon being detected as infected. However, in the absence of data on individual behaviour, we treat all cases homogeneously.

We base our analysis on the work of Stockdale et al. (2022). In particular, we use their dataset, which is available in a code repository (Irvine et al., 2023). To be precise, this dataset consists of 53 trajectories, each of which is a time series of daily infection counts. We assume that each outbreak took place at a different facility, although we only have records to support this for the multi-case outbreaks. Some cases did not have their start times recorded, which manifests in our dataset as missing values. These missing infection times have been imputed by sampling from other infection times in the same outbreak. This imputation has been repeated 100 times.

An important feature of the disease outbreaks is missing from the dataset of Stockdale et al. (2022); specifically, the infection durations. In particular, this would tell us exactly how many active cases are present at each facility on each day. Although data on infection durations is not available and likely does not exist, we can nevertheless model outbreaks using the available data by approaching this as a missing data problem.

3 Probability Model

We begin by presenting a probability model for the outbreak of COVID in a long-term care facility. Since we will only ever be analyzing a dataset whose missing infection times have been imputed, we focus only on modelling a complete trajectory of case counts. We refer to the (unobserved) dataset which contains the actual infection times for every case in our dataset as the “ideal dataset”. Although our analysis assumes fully observed case counts, recall that this represents only part of the information required to fully describe an outbreak. Specifically, we are missing the infection durations. We therefore frame our

problem as one of missing data and conclude this section by explicitly stating which parts of our model are observed and which are missing. Our presentation here is mostly informal. See Appendix A for a more explicit statement of our likelihood function and its derivatives.

We characterize an outbreak of COVID by the number of active cases on each day. This quantity depends on both the number of new cases generated each day and for how long each case remains active. We use a simple susceptible-infectious-removed (SIR) framework, where an individual is initially susceptible, may become infected over the course of the outbreak and eventually recovers or dies, whereupon they are not vulnerable to reinfection.

On a single day, we model the number of new infections generated by a single case as Poisson. Thus, the total number of new infections generated on a single day within one facility given the number of active cases on that day in that facility is also Poisson, with parameter multiplied by the number of active cases. We treat the initial number of cases in each outbreak as deterministic.

To account for the implementation of outbreak mitigation procedures, we allow the Poisson parameter to decrease over time according to an exponential damping term. That is, we model the number of new cases on day t of an outbreak as $\text{Poisson}(\phi_t)$, where $\phi_t = \phi_0 \exp(-\gamma t)$. Here, ϕ_0 is the mean daily number of infections generated by a single case in the absence of mitigation, and γ governs the effectiveness of mitigation. Note that t measures the time since the start of the current outbreak. That is, time is tracked separately for each outbreak.

It is also natural to model the daily number of new infections within a facility using the binomial distribution. For such a model, the sample size parameter would be the remaining number of susceptible people within a facility on a particular day. However, Stockdale et al. (2022) observe that there is little association between the size of a facility and the number of infections which arise over the course of an infection at that facility. In fact, the correlation is less than 5% (and is constant across imputations). As such, we proceed with our Poisson-based model, which ignores the facilities' sizes and is computationally simpler.

The duration of each case is modelled as $\text{Geometric}(\lambda)$ with support $1, 2, \dots$. Based on each case's duration, we infer the number of active cases on each day of the outbreak within a particular facility.

Putting everything together, we define the number of observed cases on day t of outbreak i given the number of active cases at the start of that day as $Y_t^i | \omega_t^i \sim \text{Poisson}(\omega_t^i \phi_t)$, where ω_t^i is the number of active cases at the beginning of day t in outbreak i . The duration of the k th new case generated on day t of outbreak i is $X_t^{i,k} \sim \text{Geometric}(\lambda)$. Each outbreak is treated as an iid observation from the distribution just described.

Recall that our dataset consists only of daily case counts, not case durations. We must therefore treat durations as missing and estimate parameters in the absence of a complete dataset. We refer to the case counts, $\{Y_t^i\}$, as the observed data, and the case durations, $\{X_t^{i,k}\}$, as the missing data. Taken together, we call $\{Y_t^i, X_t^{i,k}\}$ the complete data. We write $\theta = (\phi_0, \gamma, \lambda)$ for our parameter vector, and note that $\phi_0 \geq 0$, $\gamma \geq 0$, $0 \leq \lambda \leq 1$. It is

important to remember that missing durations are treated differently from missing infection days. The latter are handled by multiple imputation and are not modelled directly, while the former are modelled and incorporated into our fitting procedure described in the next section. We are interested in the parameter, θ , which generated the ideal dataset (see Section 2). To this end, we assume that our imputed datasets are sufficiently similar to the idea dataset that an estimate obtained from an imputed dataset can be treated as an estimate of the parameters of interest.

4 Model Fitting

In this section, we address the many challenges involved with fitting the model described in Section 3 to the data described in Section 2. Sections 4.1 and 4.2 discuss our algorithm, while Sections 4.3 and 4.4 address specific details of the implementation. Section 4.5 details our statistical inference, and finally, Section 4.6 explains our approach to handling the imputation performed by Stockdale et al. (2022).

4.1 The EM and MCEM Algorithms

The EM algorithm was introduced by Dempster et al. (1977) as a way to perform maximum likelihood estimation in the presence of missing data. This algorithm proceeds by alternating between an E-step and an M-step. In the E-step, we compute the conditional expectation of the complete data log-likelihood given the observed data. This gives us a function of θ , whose dependence is via the argument to the complete data likelihood. In the M-step, we maximize this function, thereby obtaining a new estimate of θ , which is used to evaluate the conditional expectation in the next E-step. See McLachlan and Krishnan (2008) for an in-depth introduction to the EM algorithm and some of its many extensions.

In practice, it is often difficult to compute the conditional expectation in the E-step. Instead, Wei and Tanner (1990) proposed the Monte Carlo EM (MCEM) algorithm, in which this conditional expectation is replaced by a Monte Carlo average. The observations used in this average are drawn from the conditional distribution of the missing data, given the observed data. In their seminal paper, Wei and Tanner give little guidance on how to implement the MCEM algorithm. Later, numerous authors proposed different implementations (Chan and Ledolter, 1995; Booth and Hobert, 1999; Caffo et al., 2005). We use the Ascent-Based version of the algorithm due to Caffo et al. (2005).

The M-step is essentially identical under the EM and MCEM algorithms. The only difference is the function being optimized. We carry out this optimization via the Interior-Point Newton-Raphson algorithm¹(Nocedal and Wright, 2006). The “Interior-Point” modification allows us to incorporate box-constraints on our parameters. Our implementation of the M-step makes extensive use of the `Julia` package `Optim` (Mogensen and Riseth,

2018).

4.2 The Ascent-Based MCEM Algorithm

Our implementation follows that of the Ascent-Based Monte Carlo EM (AMCEM) algorithm (Caffo et al., 2005). This method attempts to recover the ascent-property of the EM (or generalized EM) algorithm in a stochastic setting (McLachlan and Krishnan, 2008). It can be shown that iterations of the EM algorithm are guaranteed to increase the observed data likelihood, provided that the ascent-property holds. That is, provided that the objective function at our current iteration is higher when evaluated at this iteration’s parameter estimate than at the previous iteration’s estimate. Thus, precise knowledge about the EM objective function is important for guaranteeing desirable properties of the algorithm. The AMCEM algorithm frames the MCEM objective function as a noisy approximation of the EM objective function, and seeks to quantify the Monte Carlo uncertainty in this approximation.

Each iteration of AMCEM begins like the generic MCEM algorithm. We generate a Monte Carlo sample, use this sample to construct an objective function, and maximize this objective function. We then check whether our Monte Carlo sample was large enough (i.e. whether our approximation to the true EM objective function was adequate) by constructing a lower confidence bound for the improvement in the EM objective function. More formally, let $\hat{\theta}_{k-1}$ be the parameter estimate from the previous iteration, $\hat{Q}(\theta)$ be the MCEM objective function at the current iteration, and $\hat{\theta}_k = \arg \max \hat{Q}$ be the parameter estimate for our current iteration. Let $\Delta\hat{Q}(\theta_k, \theta_{k-1}) = \hat{Q}(\theta_k) - \hat{Q}(\theta_{k-1})$ be the improvement in the MCEM objective function at the current iteration, and ASE_k be the asymptotic standard error of $\Delta\hat{Q}$ based on the type of Monte Carlo sampling performed (see Equation (14) of Caffo et al., 2005). We construct a Wald-type lower confidence limit for the improvement in the EM objective function corresponding to our MCEM update,

$$\text{LCL} := \Delta\hat{Q}(\theta_k, \theta_{k-1}) - Z_{\alpha_1} ASE_k \quad (1)$$

where Z_{α_1} is the $1 - \alpha_1$ quantile of the standard normal distribution, and α_1 is set fairly high (Caffo et al. use $\alpha_1 = 0.15, 0.25$, or 0.3 , we use $\alpha_1 = 0.1$). Note that we are using LCL as a convenient measure of improvement, not as a way to do formal inference.

If the lower confidence bound just described is positive, then we can be reasonably confident that the EM ascent property holds for the current iteration, and we proceed to the next step. If the bound is negative, we augment our Monte Carlo sample (Caffo et al. suggest adding a fraction of the number of samples with which we started this iteration, e.g. $1/2$ or $1/3$; we use $1/2$), then compute a new estimate and bound. We repeat this

¹Since our objective function is an average, it is straightforward to compute its first- and second-order derivatives. Our problem is thus particularly amenable to use of the Newton-Raphson algorithm.

augmentation step until a positive bound is obtained. We begin our next iteration with a sample size equal to what we deemed sufficient at the current iteration².

The AMCEM algorithm checks for termination in a similar way. After concluding that sufficient Monte Carlo samples have been generated, we compute an upper confidence limit for the improvement in the EM objective function (using the same form as Equation (1), possibly with a different significance level). If this bound is less than some pre-specified tolerance then we terminate the algorithm. Caffo et al. refer to this method as stopping “when the change in the Q -function is too small to be easily detected”. One of their examples uses a threshold of 10^{-5} . Due to the high-dimensional nature of our dataset, we use a more liberal stopping-rule of 10^{-3} and a significance level of 0.2.

In the interest of keeping computing time manageable when repeating our analysis on many initializations and many imputations, we limit each run of the AMCEM algorithm to 10 minutes. Further investigation found that our results are not very sensitive to this choice of runtime.

In summary, we fit our model using the AMCEM algorithm, starting with a Monte Carlo sample size of 5. At each iteration, we use a 90% confidence bound to check whether the Monte Carlo size needs to be increased. If so, we augment the Monte Carlo sample by a number of new points equal to half the size we started the current iteration with. When assessing convergence, we use an 80% confidence bound, and stop when this bound is less than 10^{-3} . We also include the alternative stopping rule of terminating after the algorithm has run for ten minutes.

4.3 Importance Sampling

In order to evaluate the Monte Carlo average in the E-step of the MCEM algorithm, we must approximate expectations under the conditional distribution of the missing data given the observed data. This conditional distribution is only known up to an intractable normalizing constant. It is therefore not possible for us to sample directly from this distribution, so we instead use importance sampling. See, e.g., Section 3.3 of Robert and Casella (2004).

Our proposal distribution is based on the conditional distribution of the missing data given the past. Specifically, if the distribution of a particular duration, $X_t^{i,k}$, given the past is $\text{Geom}(\lambda)$, then our proposal distribution for $X_t^{i,k}$ is negative binomial with the same mean and standard deviation increased by a factor of 1.5. This proposal distribution behaves like an overdispersed version of the geometric. See, e.g., McCullagh and Nelder (1989) for some other examples of overdispersion. We give details about the calculation of our importance weights in Appendix B.

Furthermore, since we are unable to compute the normalizing constant for the target distribution, we normalize our importance weights to sum to 1. This “auto-normalized”, or

²Caffo et al. give an additional step here, in which we check whether the Monte Carlo size should be increased for the next iteration. In our experience, this further increase was never deemed necessary, so we omit it.

”self-normalized”, form of importance sampling causes the unknown normalizing constant to cancel and leaves us with a sampling scheme which remains consistent for the target expectation. See Elvira and Martino (2022) or Chapter 8 of Chopin and Papaspiliopoulos (2020) for more on self-normalized importance sampling.

It is a well-known feature of importance sampling that its performance depends on the variance of the weights. See, e.g., Agapiou et al. (2017). One strategy for decreasing this variance (at the expense of increasing the bias) is to truncate large weights at some threshold (Ionides, 2008). We follow the general recommendation of Ionides, and truncate weights at the threshold $\bar{w}\sqrt{M}$, where \bar{w} is the mean of the weights, and M is the size of our Monte Carlo sample. We then proceed as usual with self-normalized importance sampling, using the truncated weights. For an alternative approach to reducing the influence of large weights, see Vehtari et al. (2022). Note that truncation is performed before normalization. The truncation level is sample-specific, so a new threshold must be computed for each iteration of MCEM and whenever we augment the Monte Carlo sample within an iteration.

4.4 Multiple Starting Points

The EM algorithm, and by extension also the MCEM and AMCEM algorithms, is a local optimizer. As such, to avoid getting stuck at a point of the observed data likelihood which is locally, but not globally, optimal, we run our procedure with a range of different starting points.

A barrier to comparing results from the AMCEM algorithm across starting points is that we cannot directly evaluate the observed data likelihood. If we could, we would simply select the final estimate from the trajectory with highest likelihood. Instead, we use an identity from Chan and Ledolter (1995), which expresses an observed data likelihood ratio as the conditional expectation of the corresponding complete data likelihood ratio, given the observed data. That is, for any θ_1, θ_2 , we have

$$\frac{\mathcal{L}(\theta_1; y)}{\mathcal{L}(\theta_2; y)} = \mathbb{E}_{\theta_2} \left[\frac{\mathcal{L}_c(\theta_1; y, X)}{\mathcal{L}_c(\theta_2; y, X)} \middle| Y = y \right] \quad (2)$$

where $\mathcal{L}, \mathcal{L}_c$ are the likelihoods (not log-likelihoods) of the observed and complete data distributions respectively. See Chan and Ledolter (1995) for a derivation.

Equation (2) allows us to approximate an observed data likelihood ratio using the Monte Carlo sampler described in 4.3. Importantly, we can approximate multiple likelihood ratios using the same Monte Carlo sample, provided that all ratios have the same denominator. The approach we have described here is closely related to the Monte Carlo Maximum Likelihood method of Geyer (1994).

Note that the likelihoods which arise in our problem tend to be very small. As such, likelihood ratios can only be computed on log-scale. This poses no problem to our analysis, as we can use log-sum-exp type functions to avoid ever having to directly evaluate a likelihood ratio.

Returning now to our problem of comparing multiple AMCEM runs, we first choose a reference value for θ , say θ_* , and generate a large Monte Carlo sample under this reference value. Next, we use Equation (2) to estimate the log-likelihood ratio between θ_* and the parameter estimate from each AMCEM run. Our chosen estimator of θ is one with largest estimated likelihood ratio. Call this estimator $\hat{\theta}_d$, where the subscript d denotes that the estimator is computed on an imputed dataset.

Although we focus on the estimate with largest estimated likelihood ratio, there are several other estimates which are nearly optimal. Borrowing ideas from the cross-validation literature (see, e.g., Section 7.10 of Hastie et al., 2009), we investigate all parameter values whose estimated likelihood ratios differ from the maximum by less than one standard error. To be more precise, we first identify the maximizer of our estimated log-likelihood ratio, $\hat{\theta}_d$, where the d indicates that it is the maximizer on an imputed dataset. Next, we compute the difference between the log-likelihood ratio at $\hat{\theta}_d$ and at each other parameter estimate. Write Λ_d^s for the difference in log-likelihood ratios between $\hat{\theta}_d$ and the estimate obtained from starting point s . Next, we generate 50 new Monte Carlo samples under the reference parameter value, θ_* . On each of these new samples, we repeat the above calculations: estimate log-likelihood ratios, then take differences between $\hat{\theta}_d$ and each other parameter estimate³. Taking the standard deviation across Monte Carlo samples of each difference gives us standard errors for the original differences, $\{\Lambda_d^s\}$. Call these standard errors σ_d^s . Finally, we retain each parameter estimate whose log-likelihood difference is within one standard error of zero. That is, we keep each estimate $\hat{\theta}_d^s$ for which $\Lambda_d^s \leq \sigma_d^s$. The standard error of our log-likelihood ratio estimator is approximated by Monte Carlo (i.e., by nesting our Monte Carlo estimator inside another Monte Carlo study).

4.5 Inference

While there are multiple asymptotic standard error formulas for the estimator obtained from the classical EM algorithm (see, e.g., McLachlan and Krishnan, 2008), the one due to Louis (1982) is most amenable to use with the MCEM algorithm. Louis' standard error method is based on the formula

$$I(\theta) = \mathbb{E}_\theta [I_c(\theta; y, X) | Y = y] - \mathbb{E}_\theta [S_c(\theta)S_c(\theta)^T | Y = y] + S(\theta)S(\theta)^T \quad (3)$$

where I , I_c are the observed information matrices (i.e. negative Hessians of the log-likelihood) of the observed data and complete data distributions respectively, and S , S_c are the corresponding score vectors. See the Appendix of Louis (1982) for a derivation.

Using Equation (3), we can compute a Monte Carlo approximation to the observed information matrix of the observed data. Inverting this information matrix gives an estimate of the asymptotic covariance matrix of the observed data MLE, with which we can construct confidence sets for the parameters.

³Note that we do not choose a new maximizer on each of the new Monte Carlo samples. The goal instead is to understand the behaviour of our originally chosen maximizer.

Much interest in the world of infectious disease modelling centers around the parameter \mathcal{R}_0 , the so-called “basic reproduction number” (White, 2020). There are numerous ways to define \mathcal{R}_0 (see, e.g., Chapter 5 of Martcheva, 2015). For our context, the most natural definition is as the expected number of new infections generated by a single case over the course of their infection (in an otherwise fully susceptible population). By Wald’s Identity (see, e.g., Theorem 5.5 of Klenke, 2014), the expected number of new infections in our model is ϕ_0/λ . Since maximum likelihood estimation is invariant to re-parameterization, our MLE of \mathcal{R}_0 is $\hat{\mathcal{R}}_0 = \hat{\phi}_0/\hat{\lambda}$. Using the Delta Method (see, e.g., Chapter 3 of van der Vaart, 1998), we can estimate the asymptotic variance of $\hat{\mathcal{R}}_0$:

$$\mathbb{V}\hat{\mathcal{R}}_0 \approx \frac{\hat{\lambda}^2 \hat{\sigma}_{\phi_0}^2 - 2\hat{\lambda} \hat{\phi}_0 \hat{\sigma}_{\phi_0 \lambda} + \hat{\phi}_0^2 \hat{\sigma}_{\lambda}^2}{n \hat{\lambda}^4} \quad (4)$$

where $\hat{\sigma}_{\phi_0}^2$, $\hat{\sigma}_{\lambda}^2$ are estimated asymptotic variances of $\hat{\phi}_0$ and $\hat{\lambda}$ respectively, and $\hat{\sigma}_{\phi_0 \lambda}$ is their estimated asymptotic covariance.

4.6 Multiple Imputations

As we discussed in Section 2, our dataset has many missing values. We use the version of the dataset provided by Stockdale et al. (2022), in which missing infection start dates are sampled from other start dates within the same outbreak. This imputation is then repeated 100 times to produce 100 versions of the data. We repeat the above analysis on each of these imputations, and pool information. Our final estimate is obtained

Our final estimate of each parameter is the mean over all imputations and over all estimates within an imputation which have likelihood close to the maximum. More precisely, we first compute the mean over all estimates within an imputation with sufficiently large likelihood, then take the average of these averages. An alternative would be to instead compute a single average over all estimates of θ . We prefer the former approach as it does not privilege imputations with many successful initializations, although the difference in our final estimate is minor (neither any of the components nor the norm of the full vector change by more than 1%). We can express our estimator as follows. Let $D = 100$ be the number of imputations, S_d be the set of estimates (i.e., starting points) we retain from imputation d , and $\hat{\theta}_d^s$ be the estimate from starting point s of imputation d .

$$\hat{\theta} = \frac{1}{D} \sum_{d=1}^D \frac{1}{|S_d|} \sum_{s \in S_d} \hat{\theta}_d^s \quad (5)$$

$$= \frac{1}{D} \sum_{d=1}^D \bar{\theta}_d \quad (6)$$

where $\bar{\theta}_d$ is the mean of all retained estimates for imputation d .

To better understand the sources of variability in our problem, we decompose the error of our estimator, $\hat{\theta} - \theta$, as follows. Let $\bar{\theta}_d$ be the observed data MLE based on imputation d , and $\tilde{\theta}$ be the MLE based on our observed dataset with missing entries imputed by their true values. Recall that we refer to this as the ideal dataset (see Section 2). We call $\tilde{\theta}$ as the ideal dataset MLE. Then

$$\hat{\theta} - \theta = \frac{1}{D} \sum_{d=1}^D (\bar{\theta}_d - \tilde{\theta}_d) + \frac{1}{D} \sum_{d=1}^D (\tilde{\theta}_d - \tilde{\theta}) + (\tilde{\theta} - \theta) \quad (7)$$

$$=: \mathcal{A} + \mathcal{B} + \mathcal{C} \quad (8)$$

We refer to the three terms \mathcal{A} , \mathcal{B} and \mathcal{C} respectively as the “initialization error”, “imputation error” and “statistical error”. The initialization error measures how much our retained estimates differ from the MLE on each imputed dataset. This discrepancy represents failure of the MCEM algorithm to perfectly reproduce the observed data MLE, and is caused by both Monte Carlo variability in our algorithm and poor selection of starting point. We hope that, by filtering out only estimates with large observed data likelihood, we exclude any estimates which are converging to the wrong critical point of the observed data likelihood. This filtering should remove most of the influence of starting point from our error.

The imputation error measures how much the MLE on each imputed dataset differs from the MLE on the ideal dataset. This discrepancy represents the impact of imputation on the target of our MCEM algorithm, the MLE, and is independent of how the MLE is computed (i.e., starting point and Monte Carlo variability).

The statistical error measures how much the ideal data MLE differs from the true parameter. This discrepancy represents error due to sampling variability in an idealized version of our problem where the observed data have no missing components.

We estimate the variance of each component in our decomposition as follows. First, \mathcal{A} is the average error of retained estimates around the MLE within an imputed dataset, then averaged over the imputations. We estimate the expected square of these errors by analogy. For a fixed imputation, we compute the sample variance of the retained estimates, then we average across imputations. Next, \mathcal{B} is the average error of the imputed datasets’ MLEs from the ideal dataset MLE. We estimate the expected square of this error by the sample variance of the $\bar{\theta}_d$ ’s. That is, the sample variance of the average retained estimate within the imputations. Note that \mathcal{C} is just the difference between an MLE and its target parameter. We estimate the expected square of \mathcal{C} by estimating the inverse of the observed data information matrix⁴ for each retained estimate, then averaging within each imputation and averaging the averages across imputations. Finally, we assume that our estimator is approximately unbiased for θ and that the expected squares are additive

Table 1: Estimate for each parameter, as well as standard deviations based on three sources of variability.

Parameter	Estimate	Source of Variability		
		Statistical	Imputation	Initialization
ϕ_0 - Baseline Infectiousness	0.33	0.035	0.0064	0.0090
γ - Mitigation Effectiveness	0.053	0.0043	0.0014	0.0021
λ - Daily Recovery Probability	0.094	0.0056	0.0039	0.0059
\mathcal{R}_0 - Basic Reproduction Number	3.5	0.44	0.16	0.24

Table 2: Estimate for each parameter, as well as pooled standard deviation based on three sources of variability.

Parameter	Estimate	Pooled SD
ϕ_0 - Baseline Infectiousness	0.33	0.037
γ - Mitigation Effectiveness	0.053	0.0050
λ - Daily Recovery Probability	0.094	0.0090
\mathcal{R}_0 - Basic Reproduction Number	3.5	0.52

over our decomposition. That is, we assume that

$$\mathbb{V}\hat{\theta} \approx \mathbb{E}(\hat{\theta} - \theta)^2 \quad (9)$$

$$\approx \mathbb{E}\mathcal{A}^2 + \mathbb{E}\mathcal{B}^2 + \mathbb{E}\mathcal{C}^2 \quad (10)$$

$$\approx \widehat{\mathbb{E}\mathcal{A}^2} + \widehat{\mathbb{E}\mathcal{B}^2} + \widehat{\mathbb{E}\mathcal{C}^2} \quad (11)$$

where the terms in line (11) are the estimates described above.

5 Results

Table 1 gives our estimate of each parameter, averaged over 100 imputations. We also include the standard deviation due to each source of variability, as described in Section 4.6. Table 2 gives pooled standard deviation estimates, which are obtained by taking the square root of the sum of squared standard deviations for each estimate from Table 1.

A natural target of inference is whether mitigation is effective. This corresponds to testing the null hypothesis that $\gamma \leq 0$ (one-sided since it is unlikely that mitigation procedures increase infectivity). We use a Wald test, based on the estimate given in Table 2

⁴The word “observed” has been overloaded here. The precise matrix that we are talking about is the inverse of the negative Hessian of the log-likelihood based on the case counts in a particular imputed dataset. We are not referring to the dataset which was actually recorded (with missing infection times) and used as a basis for imputation.

and the pooled variability also given therein. Our hypothesis is rejected at the 5% level ($p < 2.2 \cdot 10^{-16}$). We therefore conclude that mitigation is effective.

Using the estimate and pooled variability given in Table 2, we construct a 95% Wald-type confidence interval for \mathcal{R}_0 . This interval ranges from 2.44 – 4.50. The value $\mathcal{R}_0 = 1$ is particularly important, since this is the threshold above which we expect outbreaks to occur. Based on our confidence interval, we have strong evidence to conclude that $\mathcal{R}_0 > 1$ (two-sided p-value $\approx 3.57 \times 10^{-11}$).

We examined results from all initializations for a few sample imputations. We found that estimates of ϕ_0 and γ were fairly consistent across initializations, but that estimates of λ varied considerably. This inconsistency in estimation of λ also leads to substantial variability in estimates of both the observed data log-likelihood and \mathcal{R}_0 . Restricting attention to only estimates with large observed data likelihood reduces this variability dramatically.

6 Conclusions and Discussion

In this final section, we present our interpretation of the results given in Section 5. We also discuss some limitations of our analysis, as well as the implications of our findings for a broader understanding of COVID-19.

6.1 Conclusions

We conducted an analysis of the transmission dynamics of SARS-CoV-2 in long-term healthcare facilities across the province of British Columbia. With access to only partial information about the outbreaks, we used missing data methodology to do maximum likelihood estimation based on the available, incomplete, data. Section 5 gives the results of our analysis. We see that mitigation does appear to be effective. This is unsurprising, and highlights the important work done by healthcare workers during the pandemic.

We also find strong evidence that the basic reproduction number, \mathcal{R}_0 , for COVID-19 is well above the critical threshold of 1. Our findings are consistent with other work from a similar period of the pandemic (see, e.g., Liu et al., 2020; Alimohamadi et al., 2020, for systematic reviews). The estimate given by Stockdale et al. (2022) is smaller than ours (their $\hat{\mathcal{R}}_0 = 2.51$, with a 90% Bayesian credible interval of 0.47 – 9.0). This is surprising, since our analysis includes numerous short outbreaks which they omitted. However, repeating our analysis without these short outbreaks gives an even larger value of $\hat{\mathcal{R}}_0 = 5.00$. Our confidence interval is also considerably more narrow than their credible interval. Some of the discrepancy between our estimates and measures of uncertainty are likely due to the incorporation of prior information about the parameters in the model of Stockdale et al.. Their uncertainty quantification is thus over a wider range of sources, whereas we focus on a smaller number of parameters and only incorporate sampling variability. Discrepancies in our analyses can also be attributed to differences in modelling assumptions. Their

model has an underlying dynamical system structure and includes an additional disease compartment, while ours is strictly individual-level and uses a very simple disease model.

Among our estimators, the statistical variability is larger than the other two sources of uncertainty for $\hat{\phi}_0$, $\hat{\gamma}$ and $\hat{\mathcal{R}}_0$. However, the initialization and statistical uncertainties are comparable for $\hat{\lambda}$. It is not surprising that λ , the daily recovery probability, is harder to estimate than the other parameters in our model. This parameter relates to the observed data only via the unobserved case durations, whereas ϕ_0 and γ are more directly associated with the case counts.

6.2 Discussion

Our analysis demonstrates a method for performing maximum likelihood analysis on a dataset with high-dimensional missing data. We use a version of the MCEM algorithm which addresses concerns about choosing the Monte Carlo size and deciding when to terminate. This method is well-suited to our problem. We present an importance sampling scheme which avoids difficult high-dimensional calculations due to the high-dimensional structure of our missing data, and give a proposal distribution which is inspired by the structure of our problem. We present a novel method for comparing results across different starting points for the MCEM algorithm, and use this method to restrict attention to ‘promising’ estimates. We give a way to estimate the standard error of our parameter estimates, and use this standard error to deduce the variability of our derived estimator for the basic reproduction number, \mathcal{R}_0 . Finally, we extend our analysis to the multiple imputations of missing infection times given by Stockdale et al. (2022), and compare the relative impacts of different sources of variability on our estimates.

There are some limitations of our analysis. For one, it is common to incorporate an “exposed” compartment into models for COVID-19. This compartment represents individuals who have been infected but who are not yet able to transmit the disease to others (Kong et al., 2022). We opted to not incorporate this feature in our model in order to minimize the amount of ‘missing information’. It is a well-known feature of the EM algorithm (and thus its derivatives) that performance depends on what fraction of the complete data information is observed (Meng and Rubin, 1994). We therefore sought to use the simplest disease structure possible while still including mechanisms for transmission and recovery.

The proposal distribution we use for importance sampling is functional, but not ideal. We often experience problems of low effective sample size (see, e.g., Agapiou et al., 2017, for a definition and discussion of effective sample size). Our results and conclusions, however, remain consistent across AMCEM runs, provided that we restrict our analysis to estimates with large observed data likelihood. We are actively working on developing an adaptive importance sampling scheme (Bugallo et al., 2017) for choosing a more effective proposal.

We base our inference on the error decomposition given in Equation (7), and the matching variance decomposition given in Equation (11). While the former decomposition is

exact, the latter decomposition of variances is assumed. Furthermore, our estimators of each term in the variance is chosen

Our analysis does not incorporate any covariates. While some information is available about the facilities, this is limited to outbreaks with more than one case. We are investigating other sources of data for this problem, and will extend our analysis as we gain access to more data.

Our work offers an alternative to the existing analysis of an important problem: that of understanding outbreaks of COVID-19 in long-term healthcare facilities. Although by now the COVID-19 pandemic has been very well studied, continued investigation helps us better prepare for future pandemics. This is especially true for the early stages of such a disease. Our analysis can therefore serve as a template for studying outbreaks of other, possibly emerging, diseases in any closed-population setting.

Appendix A Likelihood Calculations

In this appendix, we present calculations for the likelihood for our probability model. This includes the log-likelihood, score and Hessian.

See Section 3 for a description of our probability model.

A.1 Log-Likelihood

We begin by defining the relevant distributions. We model the number of new infections on a day with ω active cases and infectiousness parameter ϕ as $\text{Poisson}(\omega\phi)$, and write G_ϕ^ω for the corresponding PMF. We model the infection durations as $\text{Geom}(\lambda)$, and denote its PMF by H_λ . Let T^i and T_*^i be the day of the last new case and the final recovery respectively in outbreak i . Write I for the total number of outbreaks. The complete data

log-likelihood can be written as follows:

$$\begin{aligned} \ell_c(\theta) = & \sum_{i=1}^I \left(\left[\sum_{k=1}^{Y_1^i} \log \left(H_\lambda(X_1^{i,k}) \right) \right] + \sum_{t=2}^{T^i} \left[\log \left(G_{\phi_t}^{\omega_t^i}(Y_t^i) \right) + \sum_{k=1}^{Y_t^i} \log \left(H_\lambda(X_t^{i,k}) \right) \right] \right. \\ & \left. + \left[\sum_{t=T^i+1}^{T_*^i} \log \left(G_{\phi_t}^{\omega_t^i}(0) \right) \right] \right) \end{aligned} \quad (12)$$

$$\begin{aligned} = & \sum_{i=1}^I \left(\sum_{k=1}^{Y_1^i} \log \left(\lambda(1-\lambda)^{X_1^{i,k}-1} \right) + \sum_{t=2}^{T^i} \left[\log \left(e^{-\omega_t^i \phi_t} (\omega_t^i \phi_t)^{Y_t^i} / Y_t^i! \right) + \sum_{k=1}^{Y_t^i} \log \left(\lambda(1-\lambda)^{X_t^{i,k}-1} \right) \right] \right. \\ & \left. + \sum_{t=T^i+1}^{T_*^i} \log \left(e^{-\omega_t^i \phi_t} \right) \right) \end{aligned} \quad (13)$$

$$\begin{aligned} = & \sum_{i=1}^I \left(\sum_{t=1}^T \sum_{k=1}^{Y_t^i} \left[\log(\lambda) + (X_t^{i,k} - 1) \log(1-\lambda) \right] + \sum_{t=2}^{T^i} \left[Y_t^i \log(\omega_t^i \phi_t) - \log(Y_t^i!) \right] - \sum_{t=2}^{T_*^i} \omega_t^i \phi_t \right) \end{aligned} \quad (14)$$

$$\equiv \mathcal{C} \log \left(\frac{\lambda}{1-\lambda} \right) + \log(1-\lambda) \mathcal{D} + \sum_{i=1}^I \left[\sum_{t=2}^{T^i} Y_t^i \log(\omega_t^i \phi_t) - \sum_{t=2}^{T_*^i} \omega_t^i \phi_t \right] \quad (15)$$

where \mathcal{C} is the total number of cases observed (i.e. the sum of all Y_t^i 's), \mathcal{D} is the total number of person-days spent infectious all outbreaks (i.e. the sum of all components in X), and \equiv denotes equality up to addition of terms which do not depend on the parameters (i.e. equivalence for the purposes of maximizing the likelihood). Moving from line (13) to line (14) involves separating $G_{\phi_t}^{\omega_t^i}(Y_t^i)$ into $G_{\phi_t}^{\omega_t^i}(0)$ times terms depending on Y_t^i . We then group together all terms of the form $G_{\phi_t}^{\omega_t^i}(0)$ into a single summation.

For our next step, we substitute $\phi_t = \phi_0 e^{-\gamma t}$.

$$\ell(\theta) \equiv \mathcal{C} \log \left(\frac{\lambda}{1-\lambda} \right) + \log(1-\lambda) \mathcal{D} + \sum_{t=2}^{T^i} Y_t^i \log(\omega_t^i \phi_0 e^{-\gamma t}) - \sum_{t=2}^{T_*^i} \omega_t^i \phi_0 e^{-\gamma t} \quad (16)$$

$$= \mathcal{C} \log \left(\frac{\lambda}{1-\lambda} \right) + \log(1-\lambda) \mathcal{D} + \sum_{t=2}^{T^i} Y_t^i \log(\omega_t^i \phi_0) - \gamma \sum_{t=2}^{T^i} t Y_t^i - \phi_0 \sum_{t=2}^{T_*^i} \omega_t^i e^{-\gamma t} \quad (17)$$

$$\equiv \mathcal{C} \log \left(\frac{\lambda}{1-\lambda} \right) + \log(1-\lambda) \mathcal{D} + \sum_{t=2}^{T^i} Y_t^i \log(\phi_0) - \gamma \sum_{t=2}^{T^i} t Y_t^i - \phi_0 \sum_{t=2}^{T_*^i} \omega_t^i e^{-\gamma t} \quad (18)$$

A.2 Score

Differentiating Expression 18 gives us the score vector. We compute the derivative with respect to each parameter individually.

$$\partial_{\phi_0} \ell(\theta) = \frac{1}{\phi_0} \sum_{t=2}^{T^i} Y_t^i - \sum_{t=2}^{T_*^i} \omega_t^i e^{-\gamma t} \quad (19)$$

$$\partial_{\gamma} \ell(\theta) = - \sum_{t=2}^{T^i} t Y_t^i + \phi_0 \sum_{t=2}^{T_*^i} t \omega_t^i e^{-\gamma t} \quad (20)$$

$$\partial_{\lambda} \ell(\theta) = \frac{\mathcal{C}}{\lambda(1-\lambda)} - \frac{\mathcal{D}}{1-\lambda} = \frac{\mathcal{C} - \lambda \mathcal{D}}{\lambda(1-\lambda)} \quad (21)$$

A.3 Hessian

Differentiating each element of the score vector with respect to every parameter gives the Hessian.

$$\partial_{\phi_0}^2 \ell(\theta) = -\frac{1}{\phi_0^2} \sum_{t=2}^{T^i} Y_t^i \quad (22)$$

$$\partial_{\phi_0 \gamma} \ell(\theta) = \sum_{t=2}^{T_*^i} t \omega_t^i e^{-\gamma t} \quad (23)$$

$$\partial_{\phi_0 \lambda} \ell(\theta) = 0 \quad (24)$$

$$\partial_{\gamma}^2 \ell(\theta) = -\phi_0 \sum_{t=2}^{T_*^i} t^2 \omega_t^i e^{-\gamma t} \quad (25)$$

$$\partial_{\gamma \lambda} \ell(\theta) = 0 \quad (26)$$

$$\partial_{\lambda}^2 \ell(\theta) = \frac{\mathcal{C}(2\lambda - 1)}{\lambda^2(1-\lambda)^2} - \frac{\mathcal{D}}{(1-\lambda)^2} = \frac{\mathcal{C}(2\lambda - 1) - \lambda^2 \mathcal{D}}{\lambda^2(1-\lambda)^2} \quad (27)$$

Appendix B Importance Sampling

In this section, we describe how to compute weights for the importance sampling scheme described in Section 4.3. We address here only the calculation of raw weights (up to proportionality constant). See the main text for a discussion of post-processing (i.e. truncation

and normalization). See also Section 3 for parameter definitions, and Appendix A for definitions of the relevant densities. Write H_λ^η for the PMF of our proposal distribution. That is, a translated negative binomial distribution which starts at 1. The parameters of our proposal distribution are chosen such that its mean matches that of H_λ , and its variance is larger than that of H_λ by a factor of η^2 . More precisely, H_λ^η is the PMF of 1 plus a $\text{NegBin}(r, p)$ random variable, with $r = (1 - \lambda)/(\eta^2 - \lambda)$ and $p = \lambda/\eta^2$.

Our goal with importance sampling is to approximate the conditional distribution of the missing data given the observed data. Note that this conditional distribution is proportional to the joint distribution between the missing and observed data. Let F_θ be this joint distribution, with the observed data held fixed (i.e., treated as a function of the observed data only). We can now write our (un-normalized) importance weight, w , as

$$w(X) = \prod_{i=1}^I \prod_{t=1}^{T^i} \prod_{k=1}^{y_t^i} \frac{F_\theta(x_t^{i,k})}{H_\lambda^\eta(x_t^{i,k})} \quad (28)$$

$$= \left[\prod_{i=1}^I \prod_{t=1}^{T^i} \prod_{k=1}^{y_t^i} \frac{F_\theta(x_t^{i,k})}{H_\lambda(x_t^{i,k})} \right] \left[\prod_{i=1}^I \prod_{t=1}^{T^i} \prod_{k=1}^{y_t^i} \frac{H_\lambda(x_t^{i,k})}{H_\lambda^\eta(x_t^{i,k})} \right] \quad (29)$$

where

$$\frac{F_\theta(x_t^{i,k})}{H_\lambda(x_t^{i,k})} \propto G_{\omega_t^i}^{\phi_t}(y_t^i) \quad \text{and} \quad (30)$$

$$\frac{H_\lambda(x_t^{i,k})}{H_\lambda^\eta(x_t^{i,k})} \propto \frac{\Gamma(x_t^{i,k})}{\Gamma(x_t^{i,k} + r - 1)} (\eta^2 r)^{x_t^{i,k} - 1} \quad (31)$$

and proportionality is as a function of X . Note that we only consider t up to the day of the final observed case, because no new durations are observed thereafter.

Check for ‘‘Citation Needed’’ before publishing.

References

- S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart. Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, 32(3), 2017.
- Yousef Alimohamadi, Maryam Taghdir, and Mojtaba Sepandi. Estimate of the basic reproduction number for COVID-19: A systematic review and meta-analysis. *Journal of Preventative Medicine & Public Health*, 53(3), 2020.
- Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Vial B. Shah. Julia: a fresh approach to numerical computing. *SIAM Review*, 59(1), 2017.

- James G. Booth and James P. Hobert. Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(1), 1999.
- Mónica F. Bugallo, Víctor Elvira, Luca Martino, David Luengo, Joaquín Míguez, and Petar M. Djurić. Adaptive importance sampling: The past, the present, and the future. *IEEE Signal Processing Magazine*, 34(4), 2017.
- Brian S. Caffo, Wolfgang Jank, and Galin L. Jones. Ascent-based Monte Carlo expectation-maximization. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2), 2005.
- K. S. Chan and Johannes Ledolter. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90(429), 1995.
- Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer, 2020.
- Citation Needed.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1977.
- Víctor Elvira and Luca Martino. Advances in importance sampling, 2022.
- Charles J. Geyer. On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1), 1994.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2nd edition, 2009.
- Edward L. Ionides. Truncated importance sampling. *Journal of Computational and Graphical Statistics*, 17(2), 2008.
- Mike Irvine, Sean C. Anderson, and Jessica E. Stockdale. cr0eso, 2023. URL <https://sempwn.github.io/cr0eso/index.html>.
- Achim Klenke. *Probability Theory: A Comprehensive Course*. Springer, 2nd edition, 2014.
- Lingcai Kong, Mengwei Duan, Jin Shi, Jie Hong, Zhaorui Chang, and Zhijie Zhang. Compartmental structures used in modelling COVID-19: a scoping review. *Infectious Diseases of Poverty*, 11(72), 2022.
- Ying Liu, Albert A. Gayle, Annelies Walder-Smith, and Joacim Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2), 2020.

- Thomas A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 44(2), 1982.
- Maia Martcheva. *An Introduction to Mathematical Epidemiology*. Springer, 2015.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman & Hall/CRC, 2nd edition, 1989.
- Geoffrey J. McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*. Wiley, 2nd edition, 2008.
- Xiao-Li Meng and Donald B. Rubin. On the global and componentwise rates of convergence of the EM algorithm. *Linear Algebra and its Applications*, 199, 1994.
- Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. Optim: a mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24), 2018.
- Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2nd edition, 2006.
- Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2004.
- Jessica E. Stockdale, Sean C. Anderson, Andrew M. Edwards, Sarafa A. Iyaniwura, Nicola Mulberry, Michael C. Otterstatter, Naveed Z. Janjua, Daniel Coombs, Caroline Colijn, and Michael A. Irvine. Quantifying transmissibility of SARS-CoV-2 and impact of intervention within long-term healthcare facilities. *Royal Society Open Science*, 9, 2022.
- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. Pareto smoothed importance sampling, 2022.
- R. Vijh, C. H. Ng, M. Shirmaleki, and A. Bharmal. Factors associated with transmission of COVID-19 in long-term care facility outbreaks. *Journal of Hospital Infection*, 119, 2022.
- Greg C. G. Wei and Martin A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85(411), 1990.
- Laura F. White. Key parameters in infectious disease epidemiology. In Leonard Held, Niel Hens, Philip O’Neill, and Jacco Wallinga, editors, *Handbook of Infectious Disease Data Analysis*. Chapman & Hall/CRC, 2020.