Modeling clustered, heterogeneous transmission in a large Norovirus outbreak

Jon Zelner

March 5, 2018

Introduction

Norovirus (NoV) outbreaks are often characterized by explosive and unpredictable dynamics, as well as patterns of recrudescence in which a declining outbreak resurges. In this paper, we construct a new model of norovirus (NoV) transmission during a large outbreak to understand the role of 1) between-individual heterogeneity in infectiousness and 2) heterogeneous patterns of contact, on the overall outbreak attack rate and outbreak transmission dynamics. To do this, we re-analyzed data from a large Norovirus outbreaks first presented by Heijne et al. (1) using an individual-level, stochastic model of individual-level NoV transmission. To account for heterogeneity in individual infection potential and potential superspreading, we adapted the heterogeneous branching-process framework originally described by Lloyd-Smith et al. (2) to allow for finite numbers of susceptibles, and to accomodate contact heterogeneity.

Such contact heterogeneity, in combination with wide heterogeneity in individual infectiousness, may help to explain the "explosive, but self-limiting" character of large Norovirus outbreaks. Because transmission is often the result of vomiting events, the number of contacts an individual has around the time of illness onset is likely to be a key factor in determining his or her likelihood of either touching off a large outbreak or sustaining an ongoing one. For example, Wikswo et al. (3) documented a large NoV outbreak in which an individual who vomited during boarding of a cruise ship likely infected many passengers at one time. This individual's impact on the outbreak was likely a function of both the magnitude of the event, i.e. in terms of viral particles shed into the environment, plus the number of individuals she contacted at the time of onset. Given an equally significant vomiting event within her own room, such a passenger would likely have had a diminished epidemiologic impact.

Data

We analyze data previously published by Heijne et al. (1) from a large Norovirus outbreak during a scout jamboree in the Netherlands in 2004. Figure 1 shows the distribution of cases over time during the outbreak, and Figure 2 shows the evolution of the outbreak within the seven individual sub-camps affected during the outbreak. In this earlier analysis, Heijne et al. used a generation time approach in which the likelihood that individual i infected individual j is a function of the time lag between onset of symptoms in i and j, and assumed homogeneous patterns of contact between individuals in the different camps. The model employed in the present analysis allows us to look in more detail at the role of heterogeneous contact patterns and susceptible depletion on patterns of transmission in this outbreak, which is characterized by both explosive growth and recrudescence.

Methods

In this section, we outline a stochastic, discrete-time model of NoV transmission within and between sub-camps involved in the outbreak. This model is similar to a susceptible-exposed-infectious-recovered (SEIR) model, with key modifications to accomodate between-individual variability in infectiousness.

Lloyd-Smith et al. (2) describe an approach to modeling individual-level heterogeneity in infectiousness, in the context of *emerging* infections, to which the population is assumed to be completely susceptible and where the value of the population-level effective reproduction number is assumed to be below 1. This framework relies on the implicit assumption that large outbreaks are the result of a series of relatively improbable transmission events, rather than susceptible exhaustion. This assumption may not apply in the context of a large norovirus outbreak, such as the one described in (1) for two key reasons:

1) The average value of R may be larger than 1, indicating that susceptible depletion is a potential explanation for the end of the outbreak rather than reversion to the mean after a series of improbably infectious cases, and,
2) Outbreaks in institutional contexts, such as the scout camp in (1) or on a cruise ship (3), often reflect transmission contexts subdivided into interacting sub-populations, such as camps or decks of a ship.

In the following section, we describe an extension of this approach to the context of NoV outbreaks, specifically by modeling individual-level heterogeneity in *within-camp* infectiousness as well as the interaction between units. In our model, we denote the number of *incident* cases observed on day t in camp j to be y_{jt}. We denote the unobserved number of *infections* on day t in camp j as y^*_{jt}. These values are different because of the latent period between infection and onset of symptoms. Individuals are assumed to be infectious at the beginning of their day of onset. We denote the day of onset of disease for individual i as z_i.

Individual-level varation in infectiousness

Between-camp transmission

We assume that the average number of secondary cases generated by each case in camps other than the one they reside in is constant. We denote this *between-camp* effective reproduction number as R^B. We make this assumption for two reasons:

1) The camps are the residential contexts in this particular outbreak, and likely represent a primary locus of transmission with many opportunities for transmission in common areas, etc.
2) As the number of incident cases increases, the impact of variation is likely to decrease, as the population-level impact of a particularly infectious individual becomes more likely to be offset by, a minimally infectious case. Since the force of infection between camps is a function of the sum of all incident cases, we

Within-camp transmission

Each case is assumed to have an individual within-camp reproduction number r_i identically and individually distributed from a gamma distribution with shape parameter R^W_k and a scale parameter R^W_θ. So, $r_i \sim Gamma(R^W_k, R^W_\theta)$, and population mean value $R^W = R^W_k * R^W_\theta$. However, because within each camp multiple cases may become infectious on a given day, the values of r^W_i for each case are not individually identifiable. However, we can take advantage of the fact that the sum of n Gamma-distributed

variables is Gamma distributed with shape parameter nR_k^W to sample values r_{jt}^W indicating the daily total infectiousness across all individuals who became infectious in camp j on day t, which we can then use to estimate the daily *average* effective within-camp reproduction number for individual within that camp on day t, $\bar{r}_{jt}^W$. So, for a given camp, j:

$$r_{jt}^W = \sum_{i=1}^{y_{jt}} r_{ij}^W \mathbb{1}(t_{ij} = t)$$
$$r_{jt} \sim Gamma(y_{jt} R_k^W, R_\theta^W)$$
$$\bar{r}_{jt}^W = r_{jt}^W / y_{jt}$$

Where $\mathbb{1}(t_{ij} = t)$ is an indicator function evaluating to 1 if the time of onset in case i in camp j occurred on the day of observation, t, and zero otherwise.

Individual infectiousness as a function of time since symptom onset

We follow the approach to modeling the change in individual infectiousness after symptom onset described in (4) and previously applied to norovirus in (5). In this framework, a probability distribution is used to model the proportion of an individual's infectiousness occurring on each day following symptom onset (including the day of onset). In the current model, we allow individual infectiousness to decay according to a Geometric distribution with parameter $\gamma \in (0, 1]$, where:

$$g(x|\gamma) = \gamma(1-\gamma)^x$$

When $t \geq z_i$, $g(t - z_i)$ is the proportion of the infectiousness of individual i deposited on day t. Otherwise, when $t < z_i$, this value is equal to zero, i.e. before the individual becomes infectious. The parameter γ has a straightforward interpretation as the proportion of the individual's infectiousness deposited on the day of symptom onset, and the proportion of *remaining* infectiousness deposited each day thereafter.

Force of infection

We denote the parameter R^B to be the expected number of cases each infectious individual will create in *all other* camps except his own. Combining this with the values of within-camp infectiousness, r_{jt}^W, we can write the force of infection within each camp at each time, t, as follows:

$$\lambda_{jt} = \sum_{k \leq t} g(t-k) \left(\frac{r_{jk}}{n_j} + \sum_{i \neq j} \frac{y_{ik} R^B}{(N - n_i)} \right)$$

The risk of infection posed by each case to individuals in the other camps is a function of the size of the population outside his or her camp, i.e. $N - n_j$ for an individual residing in camp j.

Modeling the latent period

We utilize the estimates of the mean and variance of NoV incubation period duration published by Lee et al. (6) in their systematic review of incubation periods for viral gastroenteritis. Lee et al. assume that NoV

incubation periods follow a log-Normal distribution, with a median incubation time of 1.2 days (95% CI = 1.2, 1.3 days) with a dispersion factor (corresponding to the variance of the underlying normal distribution) of 1.56 (95% CI = 1.49, 1.62). In practice, this means that 95% of cases will become symptomatic within 2.5 days of infection (95% CI 2.4-2.6 days). We use a discrete approximation to this value, denoted as $\Omega(x, \mu, \sigma^2)$, where $G(x|\mu, \sigma^2)$ is the cumulative distribution function (CDF) of the log-Normal distribution, for $x \geq 1$, where x is the difference in days between infection and onset:

$$\Omega(x, \mu, \sigma^2) = G(x|\mu, \sigma^2) - G(x-1|\mu, \sigma^2)$$

Results

Within- and between-camp effective reproduction numbers

The average number of *between-camp* transmission events per case $R^W = 0.22$. (95% CI = 0.12, 0.34). By contrast, the average *within-camp* reproduction number R^W is 0.86. (95% CI = 0.41, 2.00). Summing these values gives us the average effective reproduction number $R = R^B + R^W$ = 1.07 (0.63, 2.18). However, the value of r_i across individuals is highly dispersed α = 0.03 (95% CI = 0.01, 0.05), indicating wide heterogeneity in infectiousness across individuals. Figure 5 illustrates this variation.

Figure 3 shows the evolution of r_{jt}^W across all camps over the course of the outbreak. Figure 4 shows this quantity for each camp individually. Notably, transmission within each camp is characterized by sporadic super-spreading events, rather than a time-constant rate of infectiousness.

Almost all infectiousness is concentrated on the day of onset, as indicated by the parameter estimae $\gamma =$ 0.99 (0.95, 1.00). These results echo the finding from (5), which found a strong concentration of individual infectiousness on the day of illness onset.

Between vs. within-camp transmission

The differences in the magnitude of the parameters representing within-camp infectiousness, R^W and between-camp transmission, R^B are difficult to interpret directly. This is because at any given time, there may be many more prevalent cases outside of an individual's camp than within it. As a result, even if the per-case rate of transmission is many times lower, the force of infection from outside may be significant enough that a large proportion of cases may still be acocunted for by between-camp transmission. To understand the relative contributions of within vs. between camp transmission, we use the method described by Cauchemez & Ferguson (4). To do this, we calculate the force of infection between-camp, denoted η_B and, within-camp, denoted η_W, experienced by each individual infected on each day. So:

$$\eta_j^W = \sum_{t<T} y_{jt}^* r_{jt}$$

And:

$$\eta_j^B = \zeta \sum_{t<T} y_{jt}^* \sum_{i \neq j} y_{it}$$

Where y_{it}^* is the number of new *infections* occurring on each day. Since this value is unobserved, we marginalize over the set of infection times corresponding to each incident case to estimate η_j^W and η_j^B.

For each camp, we can then estimate the proportion of cases in camp j acquired in camp j as $\tau_j = \eta_j^W/(\eta_j^W+\eta_J^B)$. We can then estimate the average prortion of infections acquired within-camp as a weighted sum across camps, $\bar{\tau} = \sum_{i=1}^{j} p_j \tau_j$, where p_j is the proportion of all incident cases occurring in camp j.

When we do this, we find that across all camps, approximately 79% of infections were acquired within-camp (95% CI = 69%,86%). Figure 6 illustrates variation in within vs. between-camp rates of infection as a function of the total incidence acocunted for by that camp. The figure illustrates a pattern compatible with the finding that most transmission is within-camp, and suggests that in the camps (e.g. 4 & 5) that had the fewest cases, seeding from outside of the camp predominated as a source of risk, whereas in those with the greatest share of cases (2 & 3), within-camp transmission predominated.

Figures & Tables

Table 1: Parameters and their values

Parameter	Definition	Estimate (95% CI)
γ	Proportion of infectiousness each day	0.99 (0.95, 1.00)
R_B	Between-camp reproduction number	0.22 (0.12, 0.34)
R_W	Average within-camp reproduction number	0.86 (0.41, 2.00)
R	Average total reproduction number ($R_W + R_B$)	1.07 (0.63, 2.18)

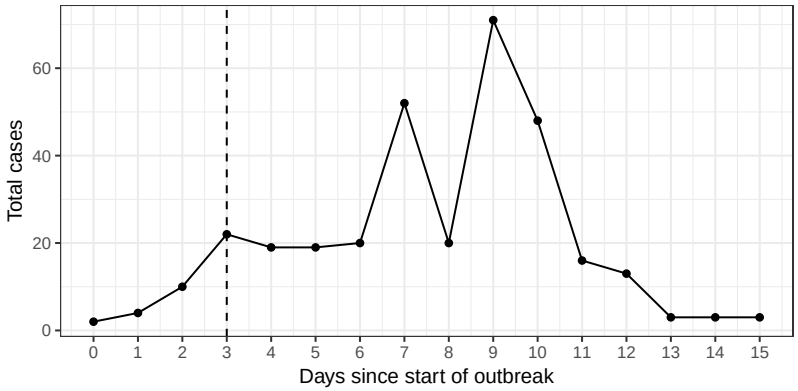

Figure 1: Number of new cases per day. Vertical line indicates beginning of hygiene interventions.

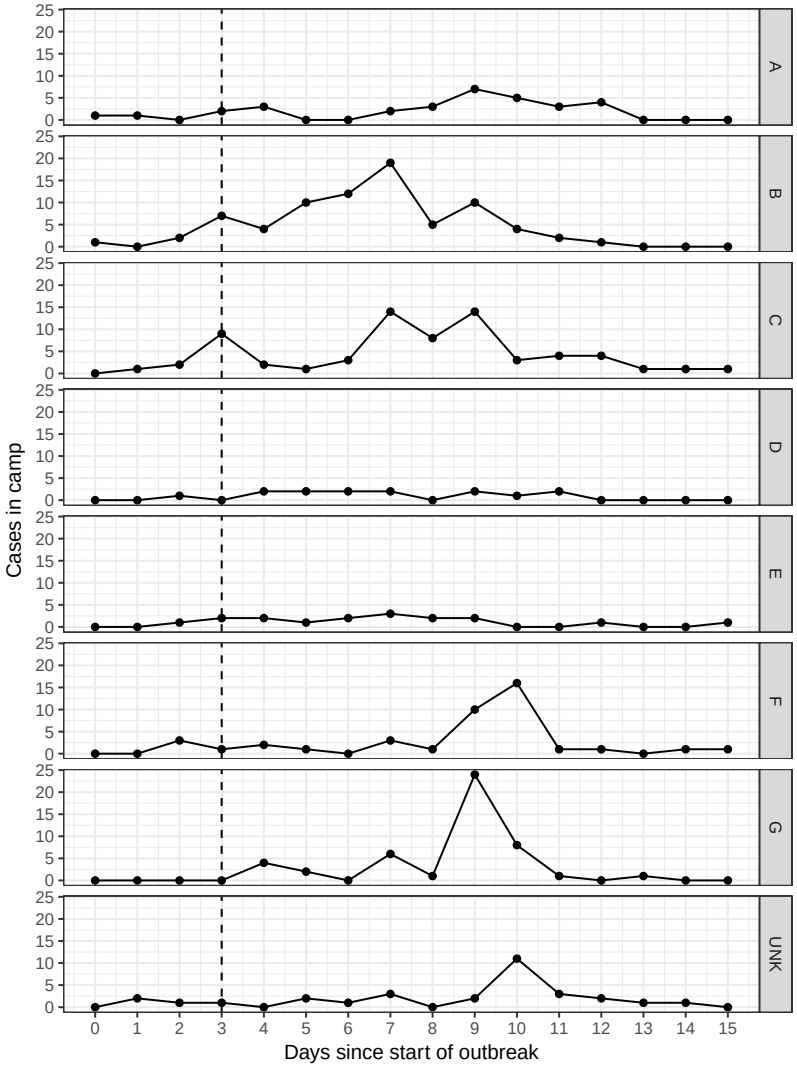

Figure 2: Number of incident cases in each camp by day of outbreak. Dotted vertical line indicates the start of outbreak interventions.

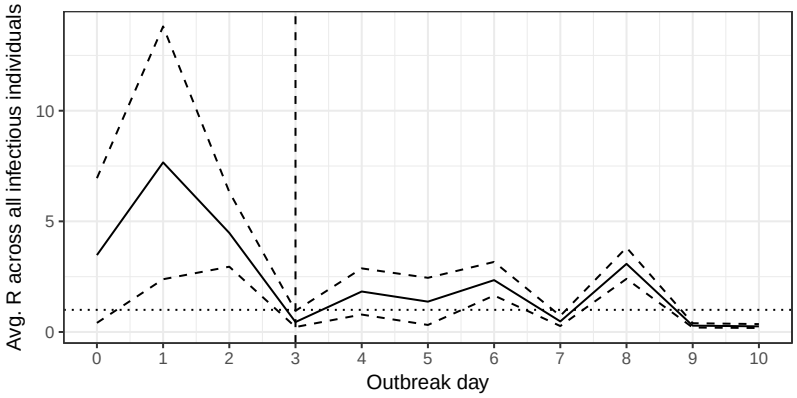

Figure 3: Estimated daily avg. reproductive number across all camps. Dashed lines indicated 95% posterior credible intervals. Dotted line indicates critical value of R = 1.

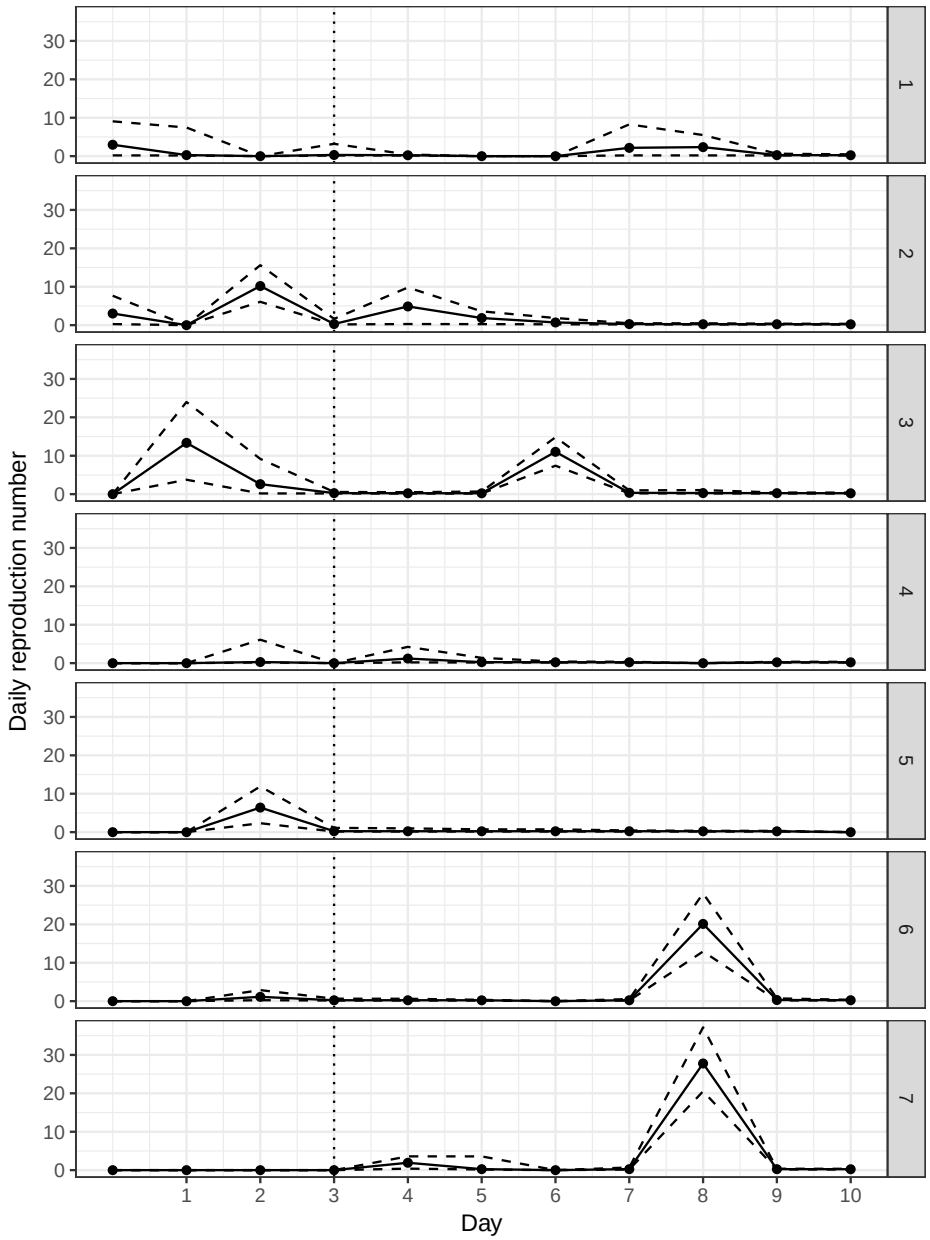

Figure 4: Estimated daily avg. reproductive number for each camp.Vertical bar indicates beginning of hygiene interventions.

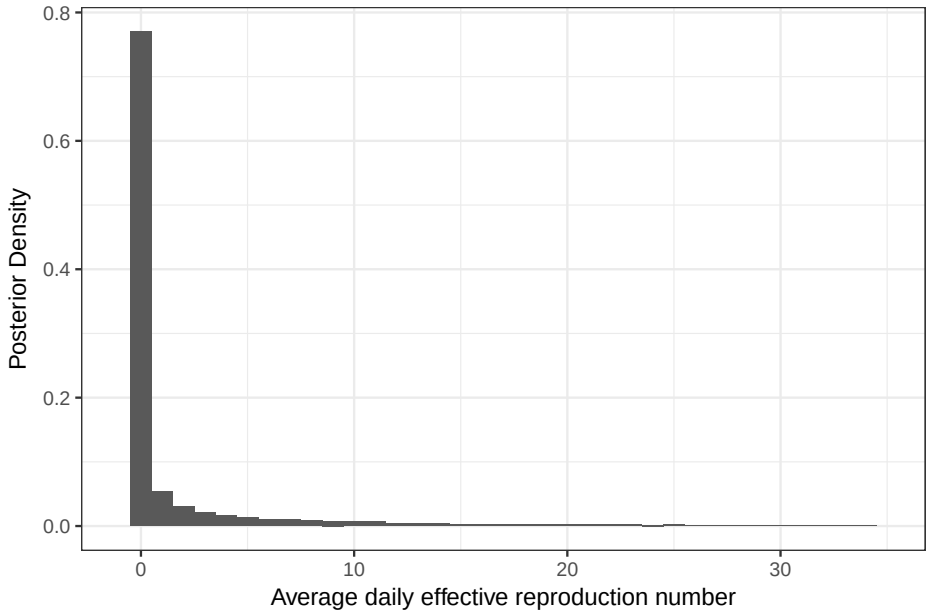

Figure 5: Marginal posterior distribution of average daily within-camp infectiousness, r_{jt}^{W}.

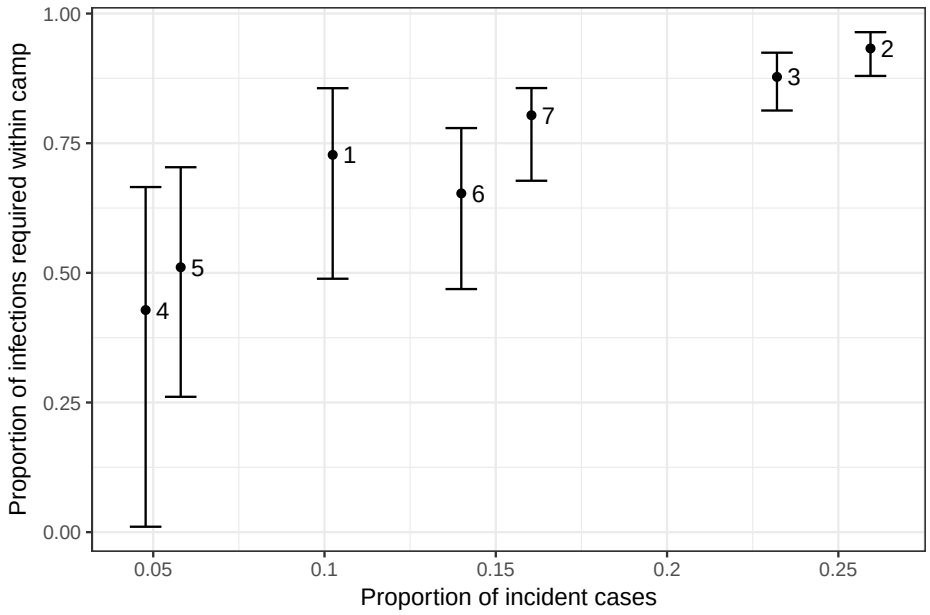

Figure 6: Proportion of infections acquired within-camp, as a function of the total proportion of incident cases within that camp. Points indicated posterior medians; vertical bars indicated 95% posterior credible intervals. Each point is labelled by the camp in Figure 2 it corresponds to.

Supplementary Methods

Likelihood

The likelihood of observing *each* of the y_{jt} cases in camp j on day $t > 1$ is a function of three components:

1) Likelihood of escaping infection from $t = 1$ until the time of infection for case i, t_i^*.
2) Likelihood of infection in camp j at $t = t_i^*$, and,
3) Likelihood of latent period of duration $z_i - t_i^*$, where z_i is the time of onset for case i.

The total system likelihood involves a fourth component:

4) Likelihood that the S_j individuals in camp j who did not develop disease during the observation period escaped infection from $t = 1$ until time T, when the jamboree ended.

The probability that an individual is infected at time t_i^* is a function of a piecewise exponential likelihood, i.e. 1 and 2 above:

$$Pr(t = t_{ij}^*) = \left(1 - e^{-\lambda_{jt}}\right) exp\left(-\sum_{i=1}^{i<t} \lambda_{ji}\right)$$

The probability that an individual has a latent period of $z_i - t_i^*$ is described by a geometric distribution with rate parameter ϵ:

$$Pr(x = z_i - t_{ij}^*) = \epsilon(1 - \epsilon)^{(z_i - t_{ij}^* - 1)}$$

For case i, the marginal probability - summing over all possible latent periods - that $t = z_i$ is:

$$Pr(z_i = t) = \sum_{j=1}^{z_i - 1} \frac{Pr(j = t_{ij}^*) Pr(x = z_i - j)}{\sum_{k=1}^{z_i - 1} Pr(x = z_i - k)}$$

Then, for all camps, j, the probability of observing y_{j2} cases on day 2, $y_j 3$ on day 3, etc, is:

$$Pr(y_{jt}) = \prod_{i=1}^{C} \prod_{t=1}^{T} y_{it} Pr(z_i = i)$$

References

1. Heijne JC, Teunis P, Morroy G, et al. Enhanced hygiene measures and norovirus transmission during an outbreak. *Emerging Infectious Diseases*. 2009;15(1):24–30.

2. Lloyd-Smith JO, Schreiber S, Kopp P, et al. Superspreading and the effect of individual variaiton on disease emergence. *Nature*. 2005;438:355–359.

3. Wikswo ME, Cortes J, Hall AM, et al. Disease transmission and passenger behaviors during a high morbidity Norovirus outbreak on a cruise ship, January 2009. *Clinical Infectious Diseases*. 2011;52:1116–1122.

4. Cauchemez S, Ferguson NM. Methods to infer transmission risk factors in complex outbreak data. *Journal of the Royal Society: Interface*. 2011;13(115).

5. Zelner JL, Lopman BA, Hall AJ, et al. Linking time-varying symptomatology and intensity of infeinfectious to patterns of Norovirus transmission. *PLOS One*. 2013;8(7):e68413.

6. Lee RM, Lessler J, Lee RA, et al. Incubation pperiod of viral gastroenteritis: A systematic review. *BMC I*. 2013;13(446).