Estimating the effective Reproductive Number (Re) for 9 Oregon SARS-Cov2 Outbreaks

Peter Banwarth

8/7/2020

Motivation and background

In order to model focal outbreaks, it is important to estimate the effective reproductive number in outbreak settings. While the estimated community R0 of 2.5-3 is useful for large scale modeling, it likely understates the Re in outbreaks simply because outbreaks are settings of higher spread.

To estimate a range of effective reproduction numbers for outbreaks, I selected 9 outbreaks recorded in Opera, a state database of communicable disease infections. These outbreaks were selected to represent a range of workplace and congregate living settings. For this analysis I excluded long term care facilities, as the disease dynamics within those facilities are likely different from general workplace/congregate living settings.

Executive Summary

The empirical effective reproductive numbers from the 9 outbreaks had a median of between 3.75 and 4.25, depending on the method. This is somewhat higher than the community R0, as expected. When running simulations of focal outbreaks, a distribution of between 3.7 and 5.0 is appropriate for generating a spread of outbreak conditions.

Methods

I accessed line lists associated with each of the nine outbreaks in the Opera Outbreaks page of Orpheus. These line lists contained all individuals associated with a given outbreak. I exported the following fields:

- Status (Confirmed, Presumptive, Suspect, Under Monitoring, No Case, Completed Monitoring)
- Epilink (Outbreak, Household, Close contact, Sporadic)
- Onset (Date of illness onset)

The resulting exports contained no individually identifiable information.

I imported the files into R and performed the following cleaning and data management steps:

1. Filtered the data to only Epilink == Outbreak. This excluded secondary cases that were not infected at the site of the outbreak.
2. Counted the total number of individuals remaining. This served as an estimate of the population of initial susceptible individuals.
3. Tabulated the dates of onset for individuals with Status %in% c(Confirmed, Presumptive). This created the epi curve for the outbreak.
4. In a few outbreaks, the date spread was so wide that the analysis threw an error. For these outbreaks I further restricted the epi curve to the "main" dates, which I chose by a judgement call. For example, in one outbreak, the final case had an onset 25 days later than the rest of the cases. I removed this case from the analysis.

I used a package in the statistical program R called "R0" and its estimation function "estimate.R" to estimate the effective reproductive number. More details can be found here: https://rdrr.io/cran/R0/ estimate.R assumes a basic SIR model, which is not generally the best model for COVID-19 spread, but is sufficient as a first approximation in an uncontrolled focal outbreak.

I ran the data through estimate.R using:

1. The generated epi.curve
2. The estimated initial susceptible population (total number of individuals with Epilink == OB)
3. The date range of the epi.curve
4. A distribution of infection probability over time. The package "R0" has a function called "generation.time" which generates this distribution.
5. estimate.R estimates Re using 5 different methods
   a. Based on fitting an uncontrolled exponential growth model to the data
   b. A maximum likelihood estimate
   c. Based on the observed attack rate
   d. A time-dependent Re, which returns Re at each point in time
   e. A sequential Bayesian process, which I have not investigated and will not be included in this report

Results

Generation Time for new illnesses

I used a gamma distribution for generation time:

generation.time("gamma",c(5,2.5))

Generation Time distribution
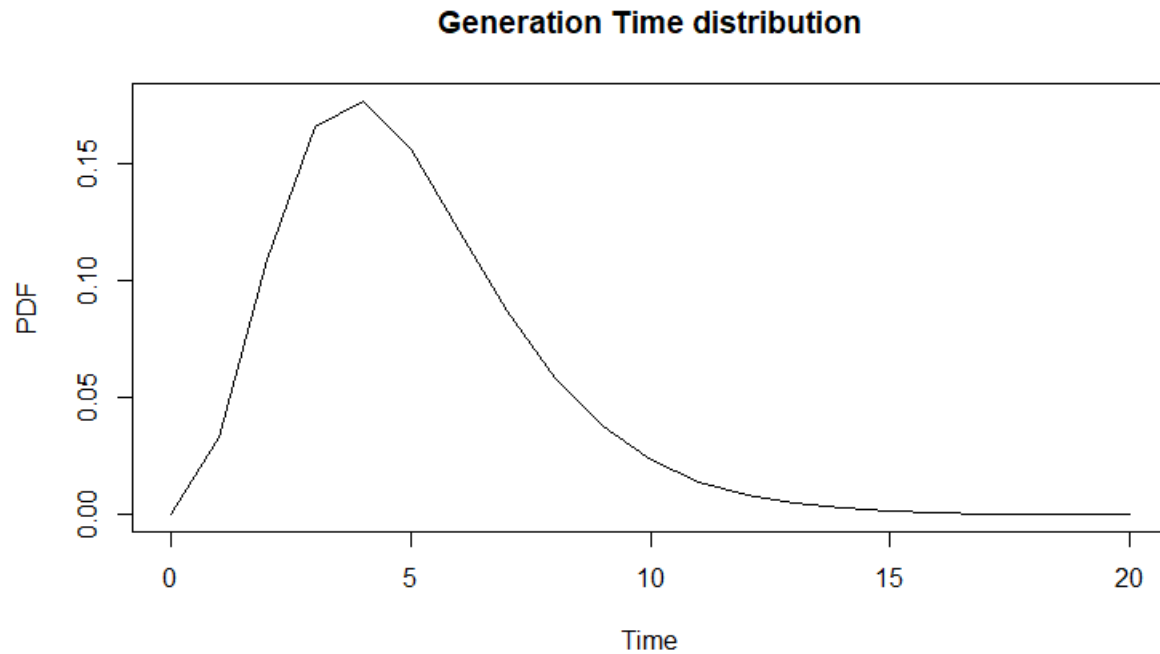
This distribution function peaks at 4 days after illness, which represents a 2-day incubation period followed by a 2-day asymptomatic period.

Outbreak 1: 2020-####

Outbreak population: 291

Total Confirmed and Presumptive: 112

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	4.1	2.8, 6.3
Maximum Likelihood	3.8	2.3, 5.7
Attack Rate	1.3	1.2, 1.3

Time Dependent (peaks at 4.5)
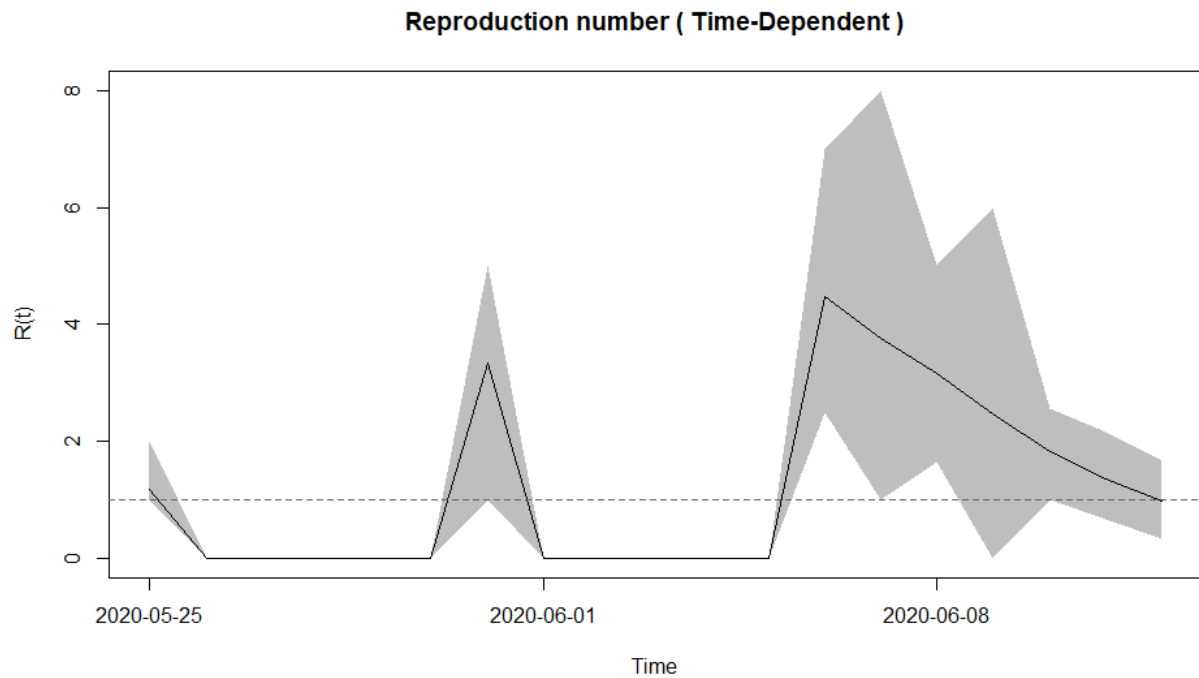
Reproduction number (Time-Dependent)

Outbreak 2: 2020-####

Outbreak population: 207

Total Confirmed and Presumptive: 170

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	10.6	7.2, 15.7
Maximum Likelihood	6.9	5.2, 9.0
Attack Rate	2.1	1.9, 2.4

Time Dependent (peaks at 5.1)
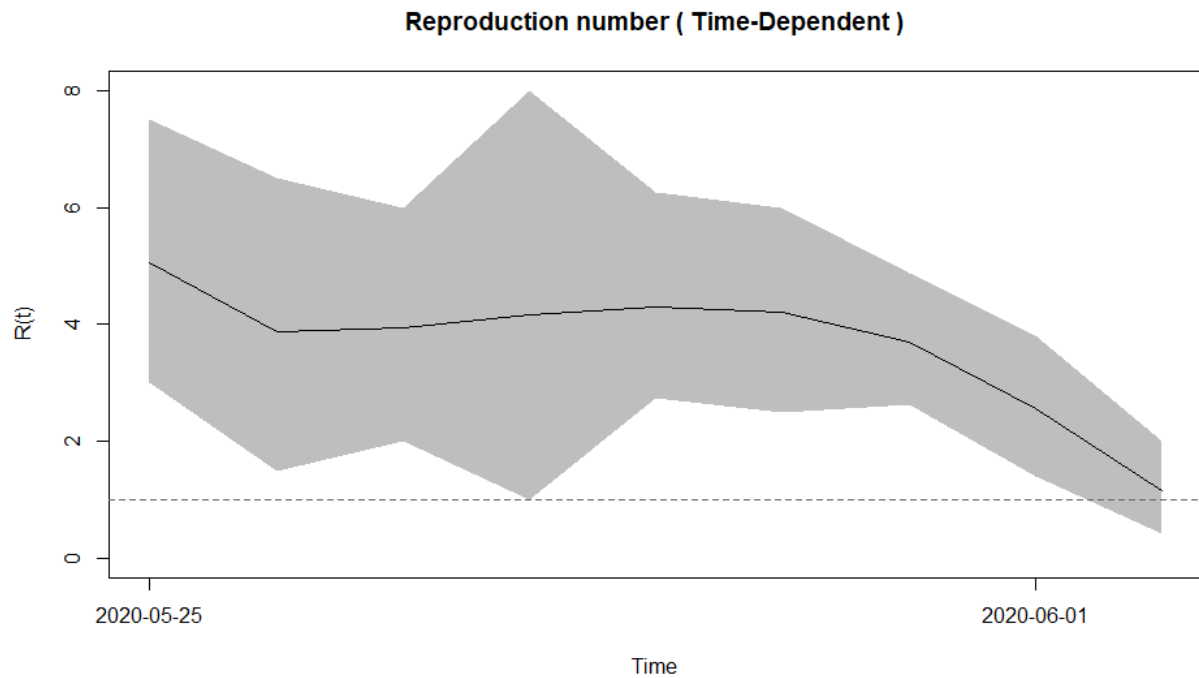
Outbreak 3: 2020-####

Outbreak population: 383

Total Confirmed and Presumptive: 184

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	2.2	1.9, 2.7
Maximum Likelihood	Could not compute	
Attack Rate	1.4	1.3, 1.4

Time Dependent (could not compute)

The data had too much variance for the Maximum Likelihood and Time Dependent methods to converge.

Outbreak 4: 2020-####

Outbreak population: 192

Total Confirmed and Presumptive: 172

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	1.4	1.3, 1.6
Maximum Likelihood	1.5	1.2, 1.8
Attack Rate	2.5	2.2, 3.0

Time dependent (peaks at 4.3)
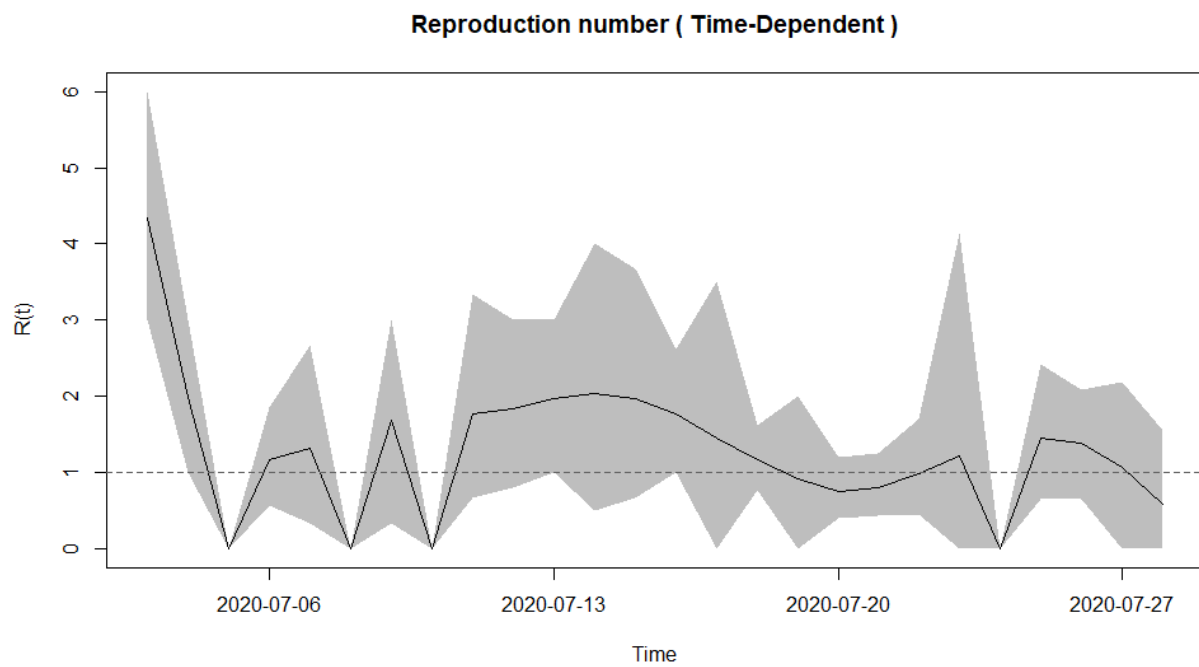
Reproduction number (Time-Dependent)

Outbreak 5: 2020-####

Outbreak population: 53

Total Confirmed and Presumptive: 44

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	1.9	0.6, 5.1
Maximum Likelihood	3.7	2.0, 6.3
Attack Rate	2.2	1.9, 3.1

Time dependent (peaks at 3.9)
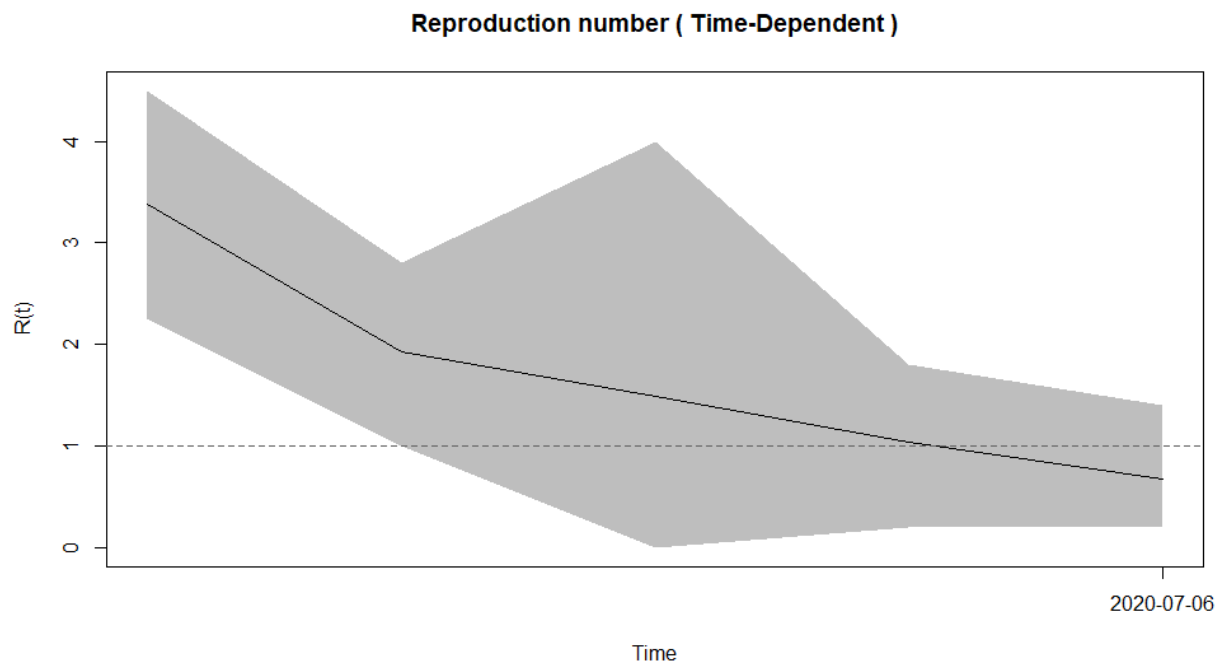
Outbreak 6: 2020-####

Outbreak population: 391

Total Confirmed and Presumptive: 275

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	6.9	5.2, 9.1
Maximum Likelihood	6.1	4.9, 7.5
Attack Rate	1.7	1.6, 1.8

Time dependent (peaks at 9.6)
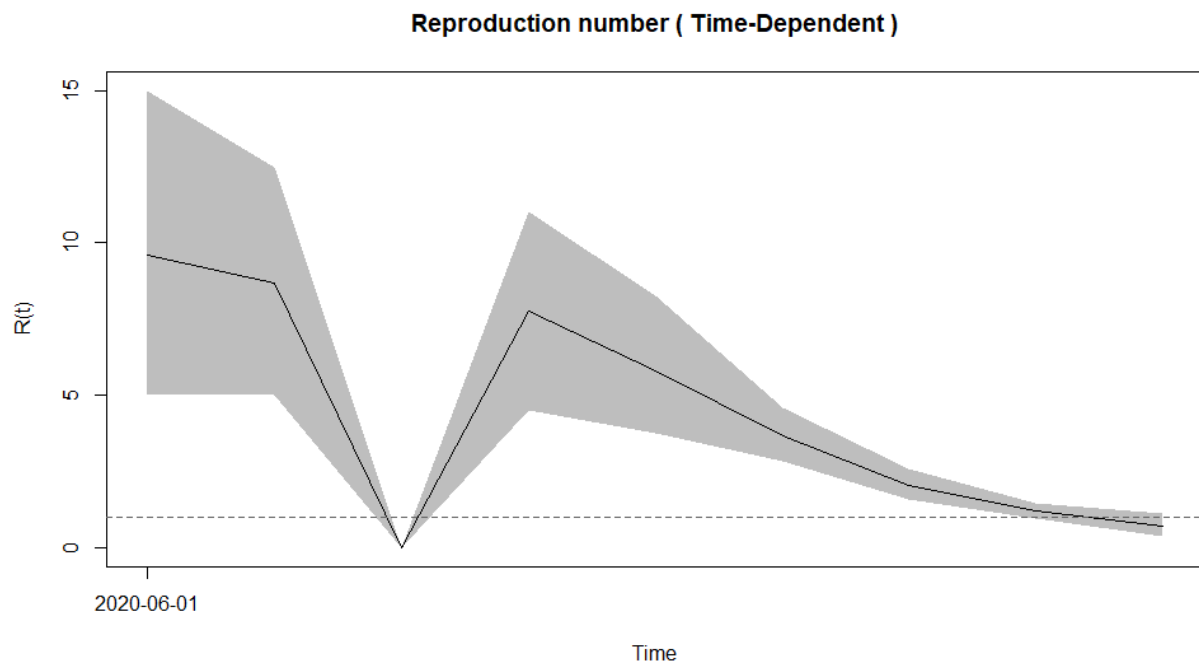
Reproduction number (Time-Dependent)

Outbreak 7: 2020-####

Outbreak population: 236

Total Confirmed and Presumptive: 164

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	4.2	3.1, 5.8
Maximum Likelihood	3.5	2.6, 4.7
Attack Rate	1.7	1.6, 1.9

Time dependent (peaks at 3.5)
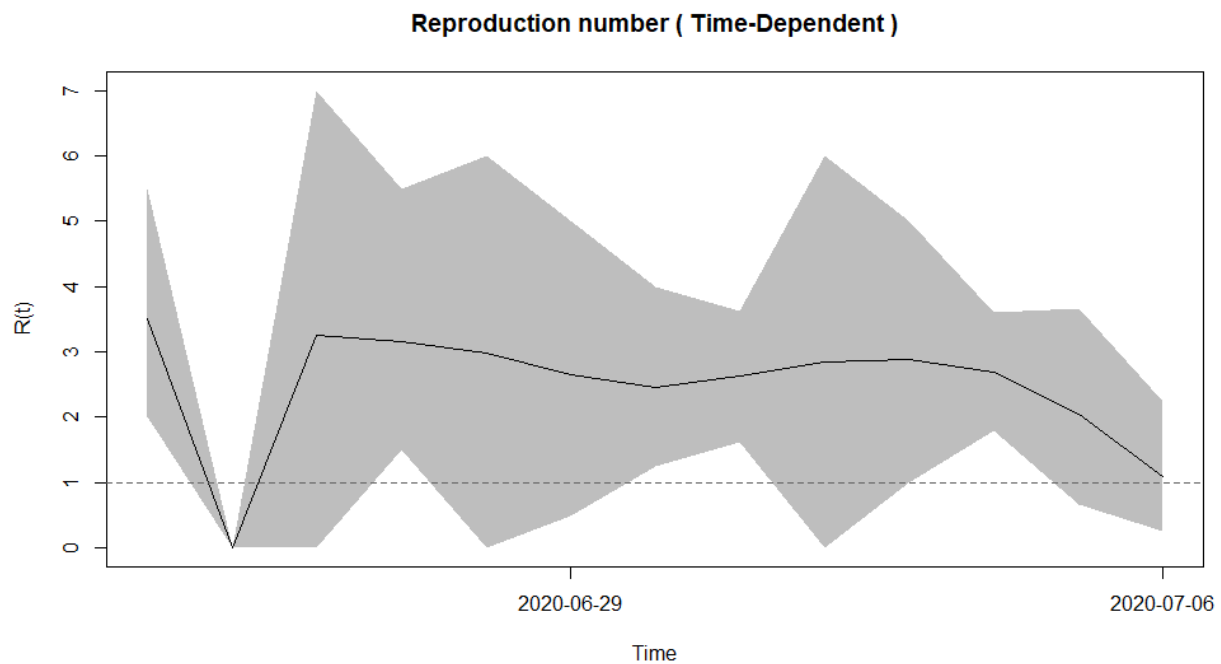
Outbreak 8: 2020-####

Outbreak population: 74

Total Confirmed and Presumptive: 40

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	5.4	1.8, 15.9
Maximum Likelihood	4.2	1.6, 8.7
Attack Rate	1.4	1.3, 1.6

Time dependent (peaks at 3.6)

Reproduction number (Time-Dependent)
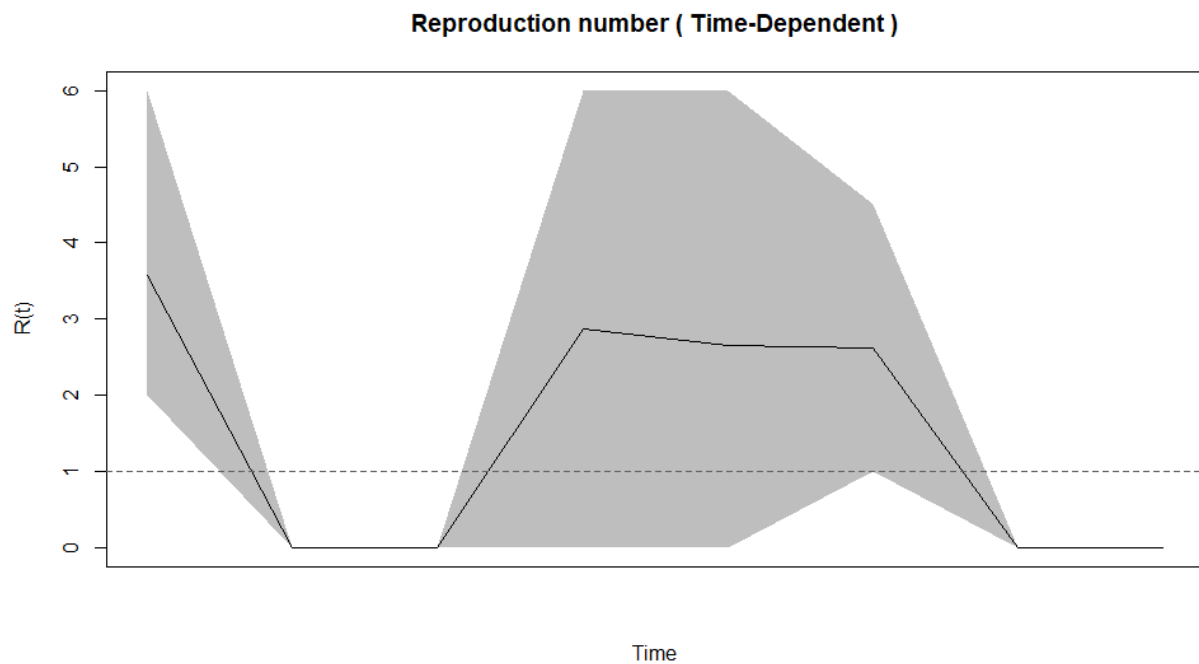
Outbreak 9: 2020-####

Outbreak population: 55

Total Confirmed and Presumptive: 49

Estimated Re:

Method	Estimate	Confidence Interval
Exponential Growth	3.5	1.9, 6.7
Maximum Likelihood	3.5	1.6, 6.5
Attack Rate	2.5	2.0, 3.7

Time dependent (peaks at 4.2)
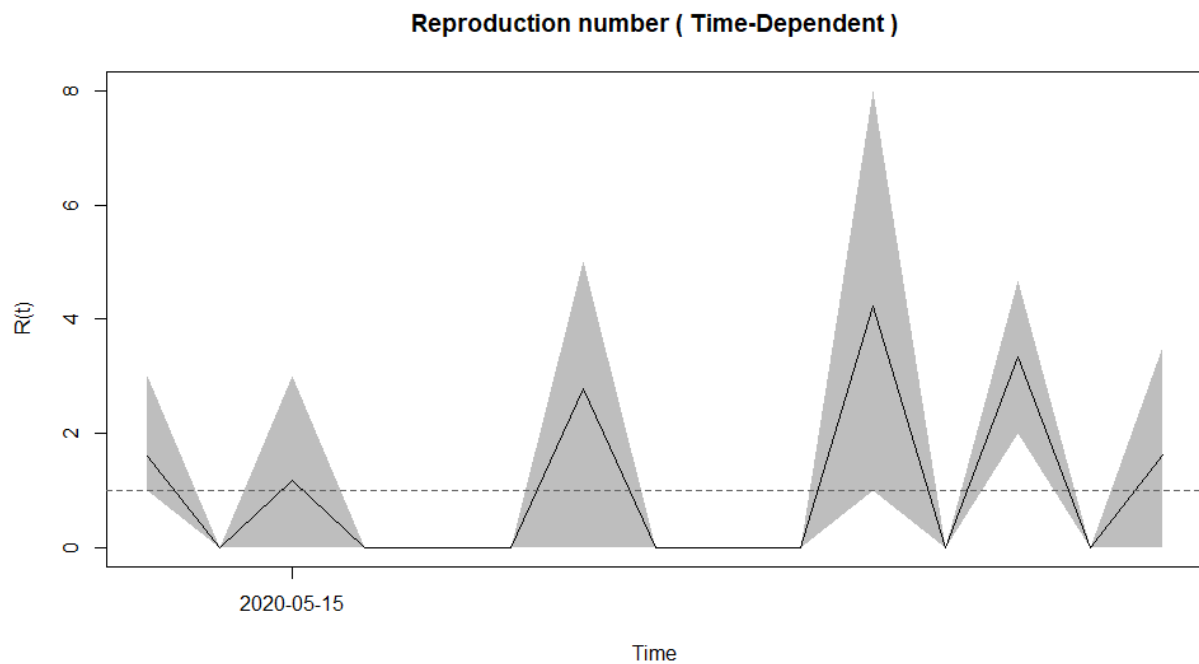
Summary of outbreaks

Outbreak	EG	ML	AR	Max TD
1	4.1	3.8	1.3	4.5
2	10.6	6.9	2.1	5.1
3	2.2		1.4	
4	1.4	1.5	2.5	4.3
5	1.9	3.7	2.2	3.9
6	6.9	6.1	1.7	9.6
7	4.2	3.5	1.7	3.5
8	5.4	4.2	1.4	3.6
9	3.5	3.5	2.5	4.2
Median	4.1	3.75	1.7	4.25
IQS	2.05, 6.15	3.5, 5.625	1.4, 2.35	3.675, 4.95

Discussion

The results obtained from this analysis are reasonable. The estimated community R0 is between 2.5 and 3.0. The median effective reproductive numbers generated from the outbreak data range between 3.75 and 4.25, depending on the method. This is higher than for community spread, as would be expected in an outbreak situation. The exception is the attack rate method, which is consistently below the empirical community R0, and therefore is not likely the appropriate method.

The COVID-Simulator script can incorporate a distribution of reproductive numbers. The IQS of the time dependent method (3.7 to 5.0) is a reasonable spread for simulations of focal outbreaks, while the IQS of the exponential growth method (2.0 to 6.2) could be used to explore fat-tail scenarios.

Limitations of this study include uncertainty of the correct onset dates; the likelihood that the estimates of the susceptible population are inaccurate (probably underestimates) due to capacity constraints during the outbreak investigation; the possibility that additional cases went undetected; and the heterogeneity of the different outbreaks. Future iterations of this study can attempt to address these limitations.