Estimating SARS-CoV-2 transmission parameters between coinciding outbreaks in a university population and the surrounding community

Matthew S. Mietchen MPH, PhD[1], Erin Clancey PhD[1], Corrin McMichael MPH, CPH[2], Eric T. Lofgren MSPH, PhD[1]

[1]Paul G. Allen School for Global Health, Washington State University, Pullman, WA

[2]Whitman County Public Health Department, Colfax, WA

Corresponding Author Email: mietchen@unc.edu

Keywords: SARS-CoV-2, COVID-19, university, sub-population, metapopulation, SEIR model, parameter estimation, likelihood function, particle filter, pMCMC

Abstract

Prior studies suggest that population heterogeneity in SARS-CoV-2 (COVID-19) transmission plays an important role in epidemic dynamics. During the fall of 2020, many US universities and the surrounding communities experienced an increase in reported incidence of SARS-CoV-2 infections, with a high disease burden among students. We explore the transmission dynamics of an outbreak of SARS-CoV-2 among university students, how it impacted the non-student population via cross-transmission, and how it could influence pandemic planning and response.

Using surveillance data of reported SARS-CoV-2 cases, we developed a two-population SEIR model to estimate transmission parameters and evaluate how these subpopulations interacted during the 2020 Fall semester. We estimated the transmission rate among the university students (β_U) and community residents (β_C), as well as the rate of cross-transmission between the two subpopulations (β_M) using particle Markov Chain Monte Carlo (pMCMC) simulation-based methods.

We found that both populations were more likely to interact with others in their population and that cross-transmission was minimal. The cross-transmission estimate (β_M) was considerably smaller [0.04×10^{-5} (95% CI: 0.00×10^{-5}, 0.15×10^{-5})] compared to the community estimate (β_C) at 2.09×10^{-5} (95% CI: 1.12×10^{-5}, 2.90×10^{-5}) and university estimate (β_U) at 27.92×10^{-5} (95% CI: 19.97×10^{-5}, 39.15×10^{-5}). The higher within population transmission rates among the university and the community (698 and 52 times higher, respectively) when compared to the cross-transmission rate, suggests that these two populations did not transmit between each other heavily, despite their geographic overlap.

During the first wave of the pandemic, two distinct epidemics occurred among two subpopulations within a relatively small US county population where university students accounted for roughly 41% of the total population. Transmission parameter estimates varied substantially with minimal or no cross-transmission between the subpopulations. Assumptions that county-level and other small populations are well-mixed during a respiratory viral pandemic should be reconsidered. More granular models reflecting overlapping subpopulations may assist with better-targeted interventions for local public health and healthcare facilities.

Introduction

SARS-CoV-2 transmission is well established to be largely airborne primarily due to aerosolized virus, which caused a worldwide pandemic beginning in late 2019. Owing to the urgent timing for policy decision-making to reduce hospitalizations and mortality, many studies, including modeling studies, being utilized early in the pandemic assumed minimal or no population-level heterogeneity, often exploring only large-scale interactions[1–7]. Geopolitical boundaries, such as states and counties, also encourage a tendency to focus on single population units due to policy decisions, as well as data collection efforts, legal jurisdictions, and access to testing and treatment programs generally governed by these same boundaries.

However, recent studies have found that various groups or subpopulations within a county, for example, have been affected by SARS-CoV-2 differently, particularly when there are large universities or industries that residents are associated with in some capacity[8–10]. The study conducted by Bharti *et al*.,2022 describes how mobile device

data showed a non-student residential population restricted movement during the pandemic and were able to show that the outbreak was primarily among students, suggesting distinct strategies for specific sub-populations should be considered for county-level outbreak management[8]. Additionally, the research by Arnold *et al*., 2022 studied SARS-CoV-2 seroprevalence among county residents, and a returning-student population were different, implying that heterogeneity in SARS-CoV-2 transmission can occur within a geographically defined population[9]. Additionally, model-based estimates such as the time-varying effective reproduction number (R_t) have been used to help governmental decision-making, such as considering when to relax various non-pharmaceutical interventions (NPIs) and were often highlighted by state officials in updates to the public. However, when a heterogenous population may be physically located in a small spatial environment such as a rural city or county, summary estimates for the whole population may be deceptive.

Like many universities, Washington State University (WSU), a large rural land-grant university located in southeastern WA, made the decision to go fully remote for all students for the Fall 2020 semester. However, many students, faculty, and staff either remained in the city or returned before the beginning of the semester. While there was no in-person instruction at the university and widespread lockdown measures were in place, a large outbreak still occurred. University testing sites, in collaboration with local public health, enabled all returning students, faculty, and staff to receive free SARS-CoV-2 (COVID-19) tests before and after the beginning of the semester. While some negative test requirements were in place, many of these relied on self-attestation. To

our knowledge, all positive student cases were initiated for contact tracing, attempted to be interviewed, and then required to isolate per university guidelines.

For members of the wider community, SARS-CoV-2 (COVID-19) testing was available throughout the city for residents at clinics, pharmacies, and non-permanent testing sites. Additionally, masking requirements and other NPI barrier methods (e.g., queue spacing, plastic dividers, etc.) were implemented throughout the city in public spaces and private businesses. Despite these mitigation efforts, the city experienced a sharp, dramatic rise in reported cases, primarily among those associated with WSU. Within the first three weeks of the semester, Whitman County, the county that WSU resides within, had one of the highest rates of reported SARS-CoV-2 in Washington State[11].

Considering a multi- or metapopulation framework allows for a higher fidelity look at subpopulation interactions, potentially identifying opportunities for public health interventions to suppress those interactions to decreases in incidence and shorter epidemic durations[12]. Metapopulation models have a rich history of describing the spatial structure of sub-populations as the driving indicator of disease dynamics and continue to support how it influences both the rate and range of pathogen dispersal and the final prevalence of disease within each population[13–16].

In this study, we use a two-population Susceptible-Exposed-Infected-Recovered (SEIR) transmission model framework to study the transmission of SARS-CoV-2 in a relatively small rural community and estimate the transmission parameters within and between two populations defined by an association with a university (or lack thereof) within a confined geographic area during the Fall of 2020. Doing so allows the

assessment of the extent to which one subpopulation might be driving the epidemic dynamics of the other subpopulation, and whether or not an intervention targeting one group (such as vaccination) may be expected to have wider population-level impacts.

Methods

Descriptive Epidemiology

Using SARS-CoV-2 infection data reported to Whitman County, Washington State, USA as well as information from Washington State University, we performed a time series analysis of weekly case counts in addition to a descriptive analysis of cases by age group, birth sex, and university affiliation between July 26, 2020, and January 2, 2021 (Morbidity and Mortality Weekly Report weeks 31 – 53). The time period for this study was chosen to capture at least a two-week timeframe before and after the beginning of the Fall semester at a large land grant university prior to a large increase in cases not previously observed in the county, while terminating before the widespread availability of vaccines for COVID-19 might alter either transmission dynamics or individual behaviors.

Due to the limitations and lack of negative test reporting, no data were available to assess SARS-CoV-2 testing positivity rates. Therefore, we assumed testing rates between those associated with the university and residents of the community were similar and not significantly different. We also assumed the availability and access to testing remained constant throughout the study period, as public health and testing capacity remained steady and, to our knowledge, did not have major fluctuations.

138 *Time-varying Reproduction Number (R_t) Estimation*

139 Due to the lack of transmission chain data or age-structured matrices and the
140 focus on sub-population transmission without assumed heterogeneity, the case
141 reproduction numbers[17] were not estimated in our study; rather we used the R_t
142 estimation method from Cori *et al.*[18]. Using the case surveillance data, we estimated the
143 instantaneous reproduction number (R_t) values for the total number of cases and the
144 sub-population-specific cases throughout the study period from Morbidity & Mortality
145 Weekly Report (MMWR) Weeks 31-53, 2020. We used the *EpiEstim* package[18,19] for
146 the R statistical computing language[20].

147 For our analysis, we assumed a parametric Serial Interval (SI) distribution with a
148 mean of 5.19 days and a standard deviation of 1.46 days from a pooled estimate in a
149 systematic review of 11 studies using a random effects model[21]. We assume the serial
150 interval distribution remained constant throughout the study period. R_t estimates were
151 adjusted for a 14-day period from the beginning of the study timeframe to ensure
152 enough case data were available for accurate estimation. The statistical framework of
153 the Cori *et al*., 2013 method has been extended to account for future incident cases
154 being imported rather than localized transmission[22]. However, we assume imported
155 cases had little to no effect due to the assumption that all university-associated
156 individuals had arrived in the city prior to the onset of the outbreak.

157

158 *Model Structure*

159 We model the transmission dynamics of SARS-CoV-2 in a closed population
160 consisting of two geographically overlapping but distinct subgroups, university-

associated individuals (denoted by the subscript U) and members of the broader community (denoted by the subscript C) with no university affiliation using a two-population compartmental model (Figure 1). The university-associated population is assumed to be primarily undergraduate students, as these represented the vast bulk of individuals tested by the university itself, while staff and faculty were primarily tested by their individual healthcare providers.

This model makes several assumptions. The total population ($N_U + N_C$) during the study period is assumed to remain constant, with no imported cases, births, or deaths. Hospitalizations and deaths had no impact on transmission and thus were not modeled directly in order to simplify the model. Once recovered, immunity to re-infection occurred throughout the length of the study. No prior immunity is assumed due to the unavailability of vaccination at this point in the pandemic and the relatively low incidence of reported SARS-CoV-2 infections prior to the large outbreak. The latency period and recovery rate were assumed to be the same for all those infected, regardless of population affiliation.

While the more simplified SEIR framework has been widely used for SARS-CoV-2 (COVID-19) transmission models, it is important to recognize that many models consider the infected state as individuals experiencing symptoms (clinical) and those who do not (subclinical) that still contribute to community transmission. While the clinical course of these two types of infections can vary widely, the absence of symptom-level information in publicly reported surveillance data, compounded with very mild symptoms experienced by many college-aged students made disambiguating these two groups untenable. We demonstrate in Supplement S1 that collapsing these two classes down

184 to a single infected “I” class effectively provides a weighted average of the two classes,
185 which accounts for the potential difference in transmission between asymptomatic and
186 symptomatic individuals at a population level.

187 We model the data collection process (i.e. weekly case counts) using the
188 negative binomial distribution. We assume that testing rate and overdispersion are
189 equivalent in both subpopulations and that testing in each subpopulation occurs
190 independently. By coupling our mathematical model of the transmission dynamics and
191 statistical model of the weekly case counts, we formulate a likelihood function (see
192 Supplement S2 and S3) that then maximize to estimate our parameters of interest.

193

194 **Parameter Estimation**

195 To further understand the disease dynamics between these two populations
196 during the COVID-19 pandemic, we used the previously described two-population
197 compartmental model to estimate the unknown transmission parameters to determine if
198 the observed epidemics are independent of one another or if there is cross-
199 transmission. Obtaining estimates of the transmission parameters requires simulation-
200 based methods which we implement using the Partially Observed Markov Process
201 (pomp) framework and R package[23,24]. We evaluated the model by simulating the
202 stochastic process of the model states to maximize the likelihood function and obtain
203 estimates of the unknown parameters listed in Table 1. Specifically, we used the particle
204 filter algorithm, as described in Algorithm 1 by King *et al., 2016,* to calculate the value of
205 the likelihood function for each set of randomly drawn unknown parameters (Table 1)
206 and the particle Markov Chain Monte Carlo algorithm (pMCMC), as described in

Algorithm 4 by King *et al*., 2016 to maximize the likelihood estimate for the observed data[23,25]. Particle Markov-chain Monte Carlo (pMCMC) was used due to its ability to estimate time-dependent hidden variables[26] and handle relatively high dimensional problems with reasonable computational demands, a challenge often found when fitting a model to surveillance data due to testing bias and underreporting. Further specifics and details of the particle filter and MCMC methods used to obtain our results are described in Supplement S4.

The initial conditions and fixed parameter values for each subpopulation used for model simulations are described in Table 1. Due to the lack of available data and small numbers for both populations, we assumed no persons were in the exposed or latent periods. Recovered individuals were defined as the number of reported cases between MMWR weeks 1 and 28 in 2020 by population affiliation. In an effort to more accurately estimate the number of infected individuals due to testing and reporting delays, we adjusted the number of reported cases, by subpopulation, during MMWR weeks 29 and 30 using SARS-CoV-2 ascertainment rates obtained for Whitman County (0.1959) and the statewide average for Washington (0.2458) in August 2020[5]. The Whitman County rate was applied to the reported cases not affiliated with the university; however, the statewide rate was used for those in the university subpopulation to account for a large number of students moving into the city prior to the Fall semester. While we believe the ascertainment rates are robust, we further adjusted the number of the infected populations to include pre-symptomatic persons that would most likely not be diagnosed or reported due to their pre-clinical stage. According to Davies *et al.,* 2020[27], the pre-symptomatic period was estimated to have a mean duration of 2.1 days out of a 5-day

230 pre-clinic/clinical duration. Therefore, we calculated the additional pre-symptomatic
231 cases to be 42% of the adjusted number of infected cases. The rate was applied to both
232 subpopulations due to the assumption of equal pre-symptomatic durations.
233

234 **Results**

235 *Descriptive Epidemiology*

236 Although the first SARS-CoV-2 (COVID-19) case in the United States was
237 reported in the state of Washington, the pandemic did not reach many rural parts of the
238 state until later in 2020. Prior to the study period (MMWR weeks 31-53; July 27, 2020 –
239 January 2, 2021), only 72 cases were reported in Whitman County, WA (pop. 47,873),
240 the county where the city and university being studied reside[28]. All reported cases
241 among city residents, including the university cases, during the first three weeks of the
242 study period (MMWR weeks 31-33) totaled 24 compared to the following three weeks
243 (MMWR weeks 34, 35, and 36), where 80, 272, and 339 cases were reported to public
244 health, respectively (Figure 2a). Additionally, the data in this figure show two distinct
245 waves, an initial large wave followed by a smaller second wave, suggesting possible
246 heterogeneities within the population.

247 Analyzing the reported case data by demographics during the study time period
248 found that cases were evenly distributed among males and females, with 77% of the
249 cases being among college-aged (18-22 years) individuals, suggesting high
250 transmission among university students. Further stratification found the initial large peak
251 to be primarily among the university population (Table 2). Examining the characteristics
252 of the initial wave, the data revealed a large epidemic among university students with no

increase in cases among the broader community. Further investigation revealed a smaller second wave of cases occurring several weeks later, primarily among the non-university-associated population, appearing around week 40, peaking between weeks 46 and 47 (Figure 2b).

Time-varying Reproduction Number (R_t) Estimation

When assuming a homogenous population using aggregated time series data, we observe one major peak with a maximum R_t value of 10.48 [8.64, 12.47] lasting 19 days with R_t >1, in addition to a second smaller peak with a maximum R_t value of 2.05 [1.62, 1.52] lasting 21 days with R_t >1 (Figure 3a). When estimating R_t for the two subpopulations, assuming no interaction or cross-infectivity, not only do we observe some of the estimates to peak at different points in time but also at very different magnitudes. While the initial R_t peak among the university-associated cases begins three days earlier compared to the initial peak among total cases, the peak remains consistent with a maximum R_t value of 11.75 [9.67, 14.01] and lasting 19 days with R_t >1, again supporting the initial peak was only among that specific population. However, the estimates for the remaining duration of the study period show large increases at different time points. Most notably, the prominent peak among community-associated cases begins 24 days before the second peak observed within the total case count time series. The magnitude of the peak is also higher, with a maximum value of 9.96 [4.96, 16.37] and lasting 17 days with R_t >1 (Figure 3b).

Parameter Estimation

We implemented the pMCMC algorithm using pomp[23,24] to estimate the unknown model transmission parameters β_U (university-university), β_C (community-community), and β_M (university-community), and additionally K and phi (Table 1). Results of the five-chain random walks are presented in Supplemental S3. The posterior distributions for the estimated model transmission parameters were: β_U: 27.92 x 10^{-5} (95% CI: 19.97 x 10^{-5}, 39.15 x 10^{-5}), β_C: 2.09 x 10^{-5} (95% CI: 1.12 x 10^{-5}, 2.90 x 10^{-5}), and β_M: 0.04 x 10^{-5} (95% CI: 0.00 x 10^{-5}, 0.15 x 10^{-5}). The full distributions are shown in Figure 4.

Both populations were more likely to interact with others in their subgroups, while the cross-transmission estimate was much smaller, or possibly zero as indicated by the credible intervals. On a relative scale, the individual transmission rates were considerably higher (β_U: 698 times and β_C: 52 times) within populations than the cross-transmission rate β_M.

Discussion

Considering the Whitman County epidemic from several different analytical standpoints, our results suggest both the university population and the wider community, while geographically proximate, experienced two separate and largely independent epidemics, rather than one epidemic simultaneously affecting both populations or an epidemic that moved from one population to the other.

Because of the larger size of the epidemic within the university population, the estimated R_t for the total population (Figure 3a) fails twice to detect distinct peaks for R_t within the community, masking the epidemic in that population almost entirely. Given the importance of R_t estimates in various governmental decisions about lifting NPIs,

missing these peaks, especially in the population more at risk for severe outcomes, has profound policy implications.

In this study, we have estimated SARS-CoV-2 transmission parameters between two populations using an SEIR modeling framework to capture concurrent epidemics. The results from estimating the three salient transmission parameters paint a similar picture as the estimates of R_t described above – namely that there appear to be two, largely independent epidemics within the community. The university population had the largest estimated transmission parameter value, roughly 13 times that of the community, but both within-population transmission parameters dwarfed the cross-population transmission parameter by 698 (University) and 52 (Community) times, respectively. Of particular note, the credible interval for β_M contained zero. The relative values of these parameter estimates suggest that transmission resulting from the mixing of these two populations played a considerably smaller role in the overall dynamics of the epidemic.

Using particle filtering MCMC, we show that two populations geographically co-occurring in a small rural town primarily spread the virus within their own groups, while cross-group transmission estimates were found to be orders of magnitude smaller than within-group transmission, suggesting separate epidemics. Additionally, comparison of the two groups, showed there was also considerable heterogeneity in both their timing and transmission rates.

Conclusion

When modeling relatively small populations within geographically defined areas, it is understandably tempting to assume homogenous mixing, especially for a disease like COVID-19 where, due to airborne transmission from aerosolized virus, brief, transient contacts in space and time between susceptible and infected individuals is sufficient to cause infection.

However, it is well documented that the social network structures of most communities vary in degrees of connectedness. In particular, communities with large universities or single, dominant employers that group individuals together through social structures and physical interactions. These groupings create highly connected subpopulations with non-static or fluctuating connections between one another, effectively creating metapopulations. Our findings suggest that during a respiratory viral pandemic, such as the COVID-19 pandemic, well-mixed population assumptions at a city or county level, even if the population is small, should be reconsidered when NPIs and other public health policies are implemented. Simplistic models (e.g., SEIR) that assume random interactions could be restructured to contain multiple populations that are weakly connected by those interactions rather than by geography are useful when social contact networks are unavailable. The computational costs of these models are often lower than individual- or network-based models, which may make them more accessible to a broader range of decision-makers. However, computational expenses could be a potential limitation depending on model complexity and other resources. While the quality of the data limits model fitting using surveillance data, model convergence, and parameter estimation are achievable.

MCMC and other simulation-based methods are dependent upon model specifications for accurate and valid results and should be implemented carefully. Improperly specified, model results may be incorrect, regardless of whether convergence is achieved. Model complexity is also considered a limitation due to convergence issues arising as the model becomes more complex. More complex models, initially explored in this study, produced convergence issues necessitating model simplification. Limitations around an unrepresentative model, for example, the inability to capture disease states such as pre-symptomatic, asymptomatic, and symptomatic for the transmission of SARS-CoV-2, may be an oversimplification of reality.

There are additional model-based limitations based on the realities of using case-data from a relatively resource constrained public health department. Parameter estimation from multiple infected classes proved difficult, as was the assessment of reliable starting values. It has been suggested that asymptomatic or subclinical case detection requires daily testing capacity and extensive serological data to improve the certainty of transmission models[29]. Subramanian *et al*., 2021, struggled with parameter identifiability however, and were leveraging data for a large city. Given the smaller, sparser data in this study, and the focus on parameter identification, the detection and testing rate of exposed persons is assumed to be equal regardless of symptomatic status. While this admittedly simplifies the overall dynamics of the model, we do not believe the lack of a clinical classification undermines the primary focus of this analysis, which is on transmission.

Lastly, an important limitation of this study is that public health policies and messaging, including legally mandated social distancing measures, were in place for the entire duration of the study period, which precludes analysis of the level to which these sub-populations mix absent preexisting interventions. We must assume that the NPIs during this time period of the pandemic amplified the separation between populations and reduced cross-transmission (indeed, this is their objective), but it is impossible to assess from our data what the cross-group transmission would be had these measures not been in place.

It is nevertheless important, anticipating future pandemics, to understand the methodological aspects of modeling both smaller rural populations and those with emergent population structures based on social and economic forces. This enables to assessment of the effectiveness of broad public health interventions, as well as estimating those that are targeted towards – or will have difficulty reaching – specific sub-populations. There was a real and justifiable concern that widespread transmission in university populations would amplify a local epidemic and then spill over into more vulnerable populations – our results suggest that this did not occur. This, in turn, offers a justification for some interventions – such as early vaccinations for the most medically vulnerable, rather than vaccinations trying to contain transmission. At the same time, it suggests that others, such as basing the suspension of some NPIs on county-level estimates of transmission, may need to be revisited.

Additional Files

Supplemental Information:

Supplement S1: Averaging of Infected Classes

Supplement S2: Two-Population Ordinary Differential Equations (ODE) Compartment Model

Supplement S3: Statistical Model and Likelihood Function

Supplement S4: Bayesian Inference

Availability of Data and Materials

Datasets and R code files are available in the **pullman_covid** GitHub repository (https://github.com/epimodels/pullman_COVID).

Acknowledgments

Funding: ETL, MSM and EC were funded by contracts 75D30121P10551, 75D30120P07911 and 75D30122C15691 from the Centers for Disease Control and Prevention, as well as RAPID 2110109 from the National Science Foundation. ETL was also supported on R35GM147013 from the National Institutes of Health.

Conflict of Interest

The authors declare they have no conflicts of interest.

References

1. Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*. 2020;20(5):553-558. doi:10.1016/S1473-3099(20)30144-4

2. Ives AR, Bozzuto C. Estimating and explaining the spread of COVID-19 at the county level in the USA. *Commun Biol*. 2021;4(1):60. doi:10.1038/s42003-020-01609-6

3. Krishnamachari B, Morris A, Zastrow D, Dsida A, Harper B, Santella AJ. The role of mask mandates, stay at home orders and school closure in curbing the COVID-19 pandemic prior to vaccination. *Am J Infect Control*. 2021;49(8):1036-1042. doi:10.1016/j.ajic.2021.02.002

4. McKee KL, Crandell IC, Hanlon AL. County-Level Social Distancing and Policy Impact in the United States: A Dynamical Systems Model. *JMIR Public Health Surveill*. 2020;6(4):e23902. doi:10.2196/23902

5. Pei S, Kandula S, Shaman J. Differential effects of intervention timing on COVID-19 spread in the United States. *Sci Adv*. 2020;6(49):eabd6370. doi:10.1126/sciadv.abd6370

6. Zou Z, Fairley CK, Shen M, et al. Critical timing and extent of public health interventions to control outbreaks dominated by SARS-CoV-2 variants in Australia: a mathematical modelling study. *Int J Infect Dis*. 2022;115:154-165. doi:10.1016/j.ijid.2021.11.024

7. Wiens K. In-person schooling and associated COVID-19 risk in the United States over spring semester 2021. *Sci Adv*. Published online 2022:12.

8. Bharti N, Lambert B, Exten C, Faust C, Ferrari M, Robinson A. Large university with high COVID-19 incidence is not associated with excess cases in non-student population. *Sci Rep*. 2022;12(1):3313. doi:10.1038/s41598-022-07155-x

9. Arnold CRK, Srinivasan S, Rodriguez S, et al. A longitudinal study of the impact of university student return to campus on the SARS-CoV-2 seroprevalence among the community members. *Sci Rep*. 2022;12(1):8586. doi:10.1038/s41598-022-12499-5

10. Schwartz EJ, Morgan M, Lapin S. Pandemic 2009 H1N1 influenza in two settings in a small community: the workplace and the university campus. *Epidemiol Infect*. 2015;143(8):1606-1609. doi:10.1017/S0950268814002684

11. Painter I, Huynh G, Ferres JML, et al. SitRep 15: COVID-19 transmission across Washington State. Published online September 8, 2020.

12. Tao Y, Hite JL, Lafferty KD, Earn DJD, Bharti N. Transient disease dynamics across ecological scales. *Theor Ecol*. 2021;14(4):625-640. doi:10.1007/s12080-021-00514-w

13. Hess G. Disease in Metapopulation Models: Implications for Conservation. *Ecology*. 1996;77(5):1617-1632. doi:10.2307/2265556

14. North AR, Godfray HCJ. The dynamics of disease in a metapopulation: The role of dispersal range. *J Theor Biol*. 2017;418:57-65. doi:10.1016/j.jtbi.2017.01.037

15. Chen S, Lanzas C, Lee C, Zenarosa GL, Arif AA, Dulin M. Metapopulation Model from Pathogen's Perspective: A Versatile Framework to Quantify Pathogen Transfer and Circulation between Environment and Hosts. *Sci Rep*. 2019;9(1):1694. doi:10.1038/s41598-018-37938-0

16. Citron DT, Guerra CA, Dolgert AJ, et al. Comparing metapopulation dynamics of infectious diseases under different models of human movement. *Proc Natl Acad Sci*. 2021;118(18):e2007488118. doi:10.1073/pnas.2007488118

17. Wallinga J. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *Am J Epidemiol*. 2004;160(6):509-516. doi:10.1093/aje/kwh255

18. Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am J Epidemiol*. 2013;178(9):1505-1512. doi:10.1093/aje/kwt133

19. Cori A. *EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves*.; 2021. https://CRAN.R-project.org/package=EpiEstim

20. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. https://www.R-project.org/

21. Rai B, Shukla A, Dwivedi LK. Estimates of serial interval for COVID-19: A systematic review and meta-analysis. *Clin Epidemiol Glob Health*. 2021;9:157-161. doi:10.1016/j.cegh.2020.08.007

22. Thompson RN, Stockwin JE, van Gaalen RD, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*. 2019;29:100356. doi:10.1016/j.epidem.2019.100356

23. King AA, Nguyen D, Ionides EL. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *J Stat Softw*. 2016;69(12):1-43. doi:10.18637/jss.v069.i12

24. King AA, Ionides EL, Bretó CM, et al. *Pomp: Statistical Inference for Partially Observed Markov Processes*.; 2022. https://kingaa.github.io/pomp/

25. Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc Natl Acad Sci*. 2015;112(3):719-724. doi:10.1073/pnas.1410597112

26. Endo A, van Leeuwen E, Baguelin M. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics*. 2019;29:100363. doi:10.1016/j.epidem.2019.100363

27. Davies NG, Klepac P, Liu Y, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*. 2020;26(8):1205-1211. doi:10.1038/s41591-020-0962-9

28. *Washington State Department of Health COVID-19 Data Dashboard*. Washington State Department of Health; 2022. https://doh.wa.gov/emergencies/covid-19/data-dashboard

29. Subramanian R, He Q, Pascual M. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc Natl Acad Sci*. 2021;118(9):e2019716118. doi:10.1073/pnas.2019716118

30. Hakki S, Zhou J, Jonnerby J, et al. Onset and window of SARS-CoV-2 infectiousness and temporal correlation with symptom onset: a prospective, longitudinal, community cohort study. *Lancet Respir Med*. 2022;10(11):1061-1073. doi:10.1016/S2213-2600(22)00226-0

31. *Total Student Enrollment - Washington State University*. Washington State University; 2022. https://ir.wsu.edu/total-student-enrollment/

32. *Washington State University Environmental Health and Safety*.; 2022. https://ehs.wsu.edu/public-health/contact-tracing/

33. *State of Washington 2021 Population Trends*.; 2022. https://ofm.wa.gov/sites/default/files/public/dataresearch/pop/april1/ofm_april1__poptrends.pdf

34. *Whitman County Public Health Department*.; 2021. https://whitmancountypublichealth.org/community-health/disease-prevention/covid-19/

Tables

Three parameters define SARS-CoV-2 viral transmission rates among and between these subgroups; β_U (university-university), β_C (community-community), and β_M (university-community). Exposed populations are within the latency period (α) or the time period between exposure and infectiousness, and those recovered are no longer infectious and leave the infectious period at a rate of γ or the recovery rate. Two additional parameters were estimated, ϕ or the testing probability and κ which is a variance constant (Table 1).

Table 1. Model Parameters and Initial Conditions

Parameter	Parameter Description	Value	Source(s)
β_U	Transmission Rate between university-associated individuals	Unknown	Estimated
β_C	Transmission Rate between individuals in the community	Unknown	Estimated
β_M	Cross-Transmission Rate between the university and community populations	Unknown	Estimated
α	Latency Period or the time between a person being exposed and becoming infectious with a pathogen	3.59 $days^{-1}$	[5]
γ	Recovery Rate or the duration of infectiousness	5 $days^{-1}$	[27,30]
ϕ	Testing probability or the proportion of those who are infected and eligible for testing (parameter is required by the model)	Unknown	Estimated
κ	Variance constant (parameter is required by the model)	Unknown	Estimated
Total University Population (N_U)	Estimated number of enrolled undergraduate students at WSU (Pullman Campus) for the 2020 Fall semester living on or near campus	14,254	[31]
Susceptible (S_U)	Number of susceptible university-associated individuals [$S_U = N_U - (E_U + I_U + R_U)$]	14,158	Calculated
Exposed (E_U)	Number of individuals associated with the university known to be exposed to SARS-CoV-2	0	Assumed
Infected (I_U)	Estimated number of infected individuals associated with the university in the two weeks prior to the study period (2020 MMWR Week 29 & 30)	81	[5,27,32]

Recovered (R_U)	Number of reported SARS-CoV-2 cases that were associated with the university prior to the study period, who are no longer infectious (2020 MMWR Week 1-28)	15	[32]
Total Community Population (N_C)	Estimnated number of non-student residents in the city of Pullman, WA	20,785	[33]
Susceptible (S_C)	Number of susceptible individuals not associated with the university [$S_C = N_C – (E_C + I_C + R_C)$]	20,642	Calculated
Exposed (E_C)	Number of individuals not associated with the university known to be exposed to SARS-CoV-2	0	Assumed
Infected (I_C)	Estimated number of infected individuals not-associated with the university in the two weeks prior to the study period (2020 MMWR Week 29 & 30)	124	[5,27,34]
Recovered (R_C)	Number of reported SARS-CoV-2 cases among the residents of Pullman, WA who were not university-associated prior to the study period, who are no longer infectious (2020 MMWR Week 1-28)	19	[34]

548 Table 2. Reported SARS-CoV-2 Infections from 07/26/2020 – 01/02/2021 by Age
549 Group, Sex, and Sub-population Demographic, Pullman, WA
550

University Sub-population

Age Group	Female (%) †	Male (%) †	Total (%) †
0-17	1 (0.1%)	0 (0%)	1 (0.1%)
18-22	740 (92.4%)	705 (87.8%)	1,445 (90.1%)
23-32	42 (5.2%)	73 (9.1%)	115 (7.2%)
33-39	7 (0.9%)	9 (1.1%)	16 (1.0%)
40-59	10 (1.3%)	12 (1.5%)	22 (1.4%)
60 and above	1 (0.1%)	4 (0.5%)	5 (0.3%)
Total (%) *	**801 (49.9%)**	**803 (50.1%)**	**1,604**

Community Sub-population

Age Group	Female (%) †	Male (%) †	Total (%) †
0-17	22 (9.5%)	14 (5.9%)	36 (7.6%)
18-22	50 (21.6%)	47 (19.7%)	97 (20.6%)
23-32	55 (23.7%)	70 (29.3%)	125 (26.5%)
33-39	22 (9.5%)	20 (8.4%)	42 (8.9%)
40-59	47 (20.3%)	51 (21.3%)	98 (20.8%)
60 and above	36 (15.5%)	37 (15.5%)	73 (15.5%)
Total (%) *	**232 (49.3%)**	**239 (50.7%)**	**471**

551 † Percentages across Age groups category

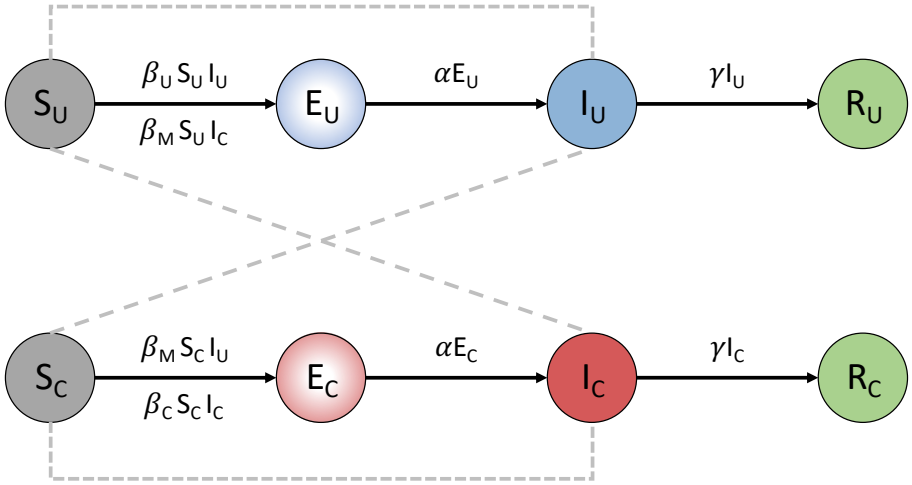
552 *Percentage across Sex category

553 Note: Percentage may not add up evenly due to rounding.

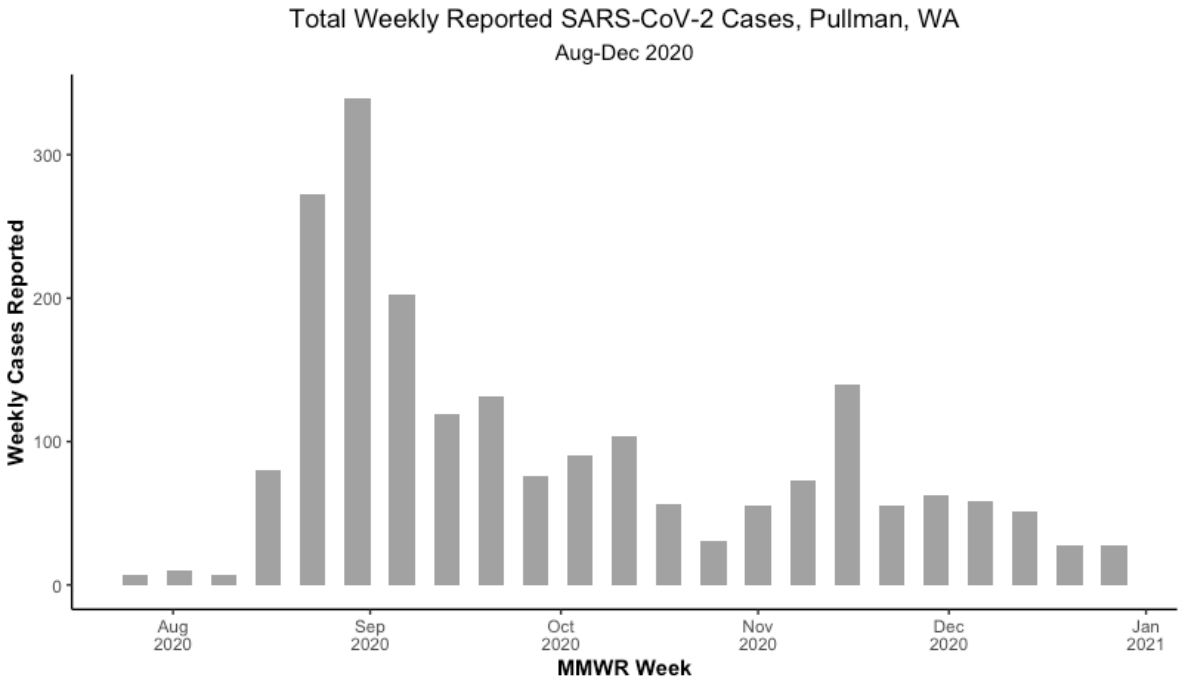
554
555
556
557
558

Figures

Figure 1. Two-population SEIR Model Schematic of the University (U) and Community (C) population interactions



579 Figure 2a. Reported SARS-CoV-2 epidemiological curve by MMWR week



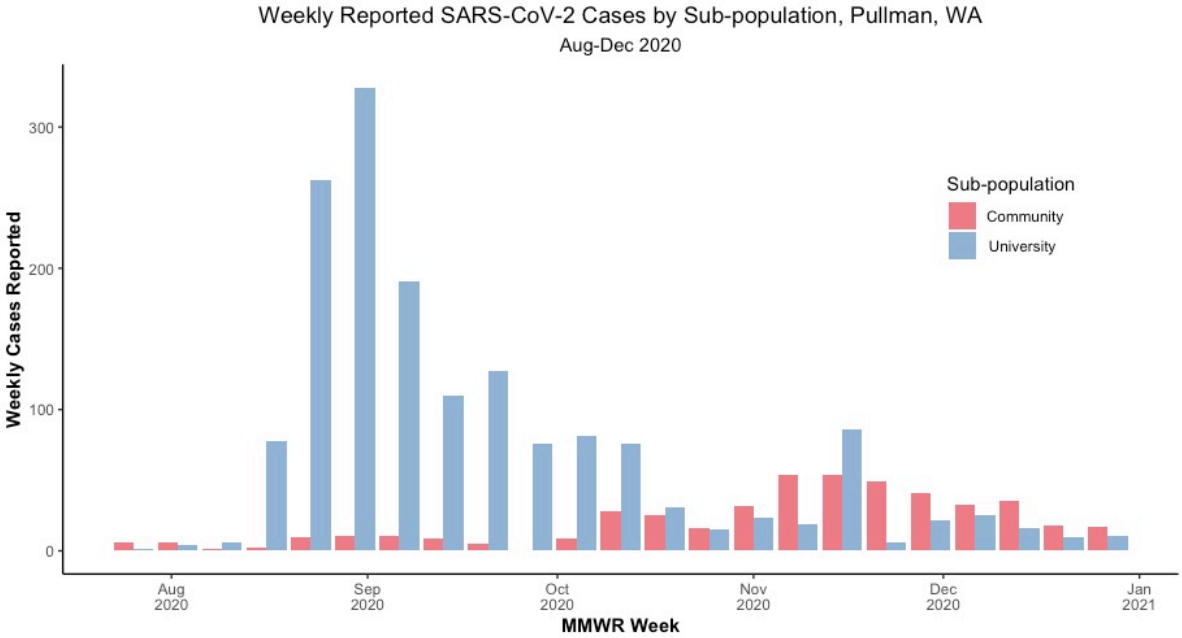
580

581

582

583 Figure 2b. Reported SARS-CoV-2 epidemiological curve by MMWR week and

584 subpopulation



585

586

Figure 3a. R_t estimates of Total SARS-CoV-2 Reported Cases

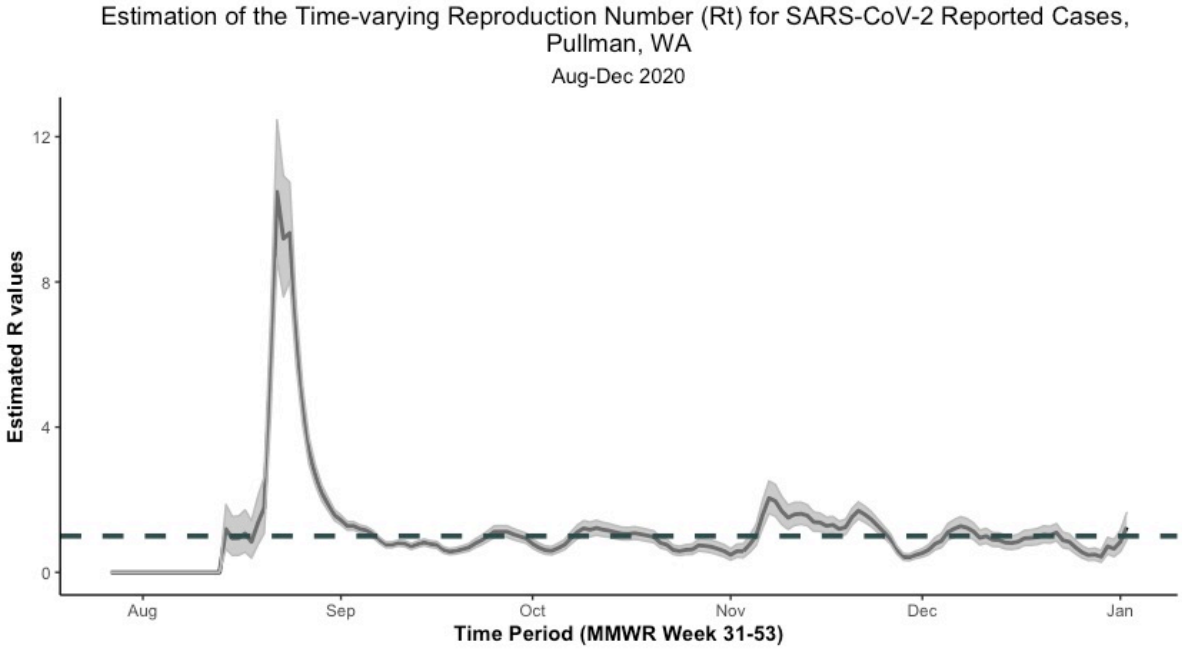


Figure 3b. R_t estimates of SARS-CoV-2 Reported Cases by Sub-Population

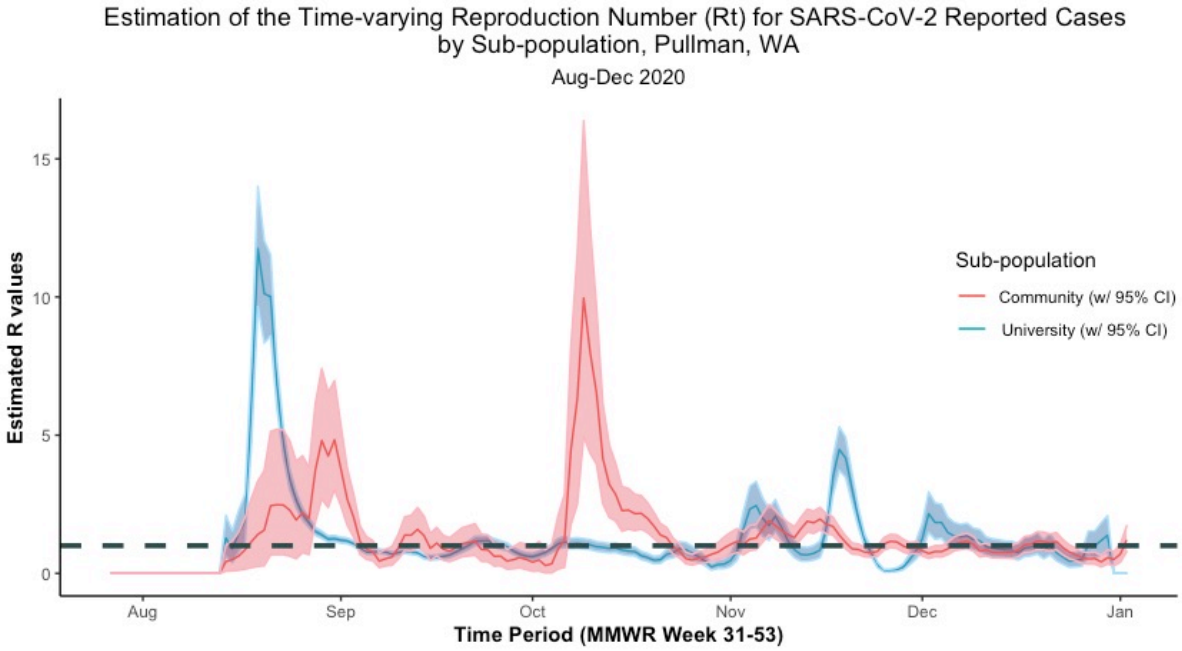
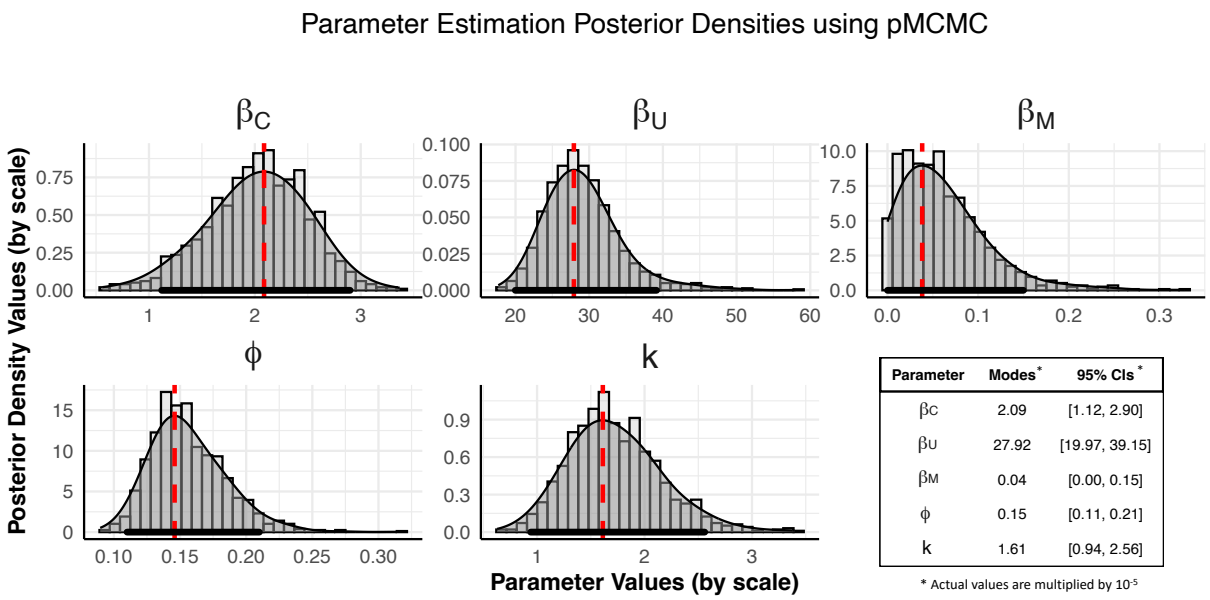


Figure 4. Posterior distributions of Processed pMCMC for Parameter Estimation. The red vertical dashed line represents the modal posterior estimate, with 95% credible intervals shown as the solid black line on the x-axis for each estimate.



Parameter	Modes*	95% CIs*
β_C	2.09	[1.12, 2.90]
β_U	27.92	[19.97, 39.15]
β_M	0.04	[0.00, 0.15]
ϕ	0.15	[0.11, 0.21]
k	1.61	[0.94, 2.56]

* Actual values are multiplied by 10^{-5}