

MODELING HEALTHCARE WORKER BEHAVIORS FROM LOCATOR TRACES

by

Eldaleona Odole

A thesis submitted in partial fulfillment
of the requirements for the Bachelor of Science
degree in Data Science Honors in the
College of Liberal Arts and Sciences at
The University of Iowa

December 2022

Thesis Committee: Alberto Maria-Segre

ACKNOWLEDGMENTS

I would like to thank the University of Iowa Computer Science Department as well as Dr. Alberto Segre for his mentorship throughout the project. I would also like to thank Dan Sewell, Rhonda DeCook and Phil Polgreen for their roles consulting on this project. Part of this research was conducted as part of the department's "Computing for Health and Well-Being REU program" whose funding was provided by the National Science Foundation (PIs Liza Kleiman and Octav Chipara) under award 2050526. Accordingly I would like to thank the other REU program participants working under CompEpi, Maren Gingerich, Vivian Nguyen, Jackie Valencia, and Amanda Pereira. Finally, this project would not be possible without the funding provided by the CDC MInD Healthcare group under cooperative agreement U01-CK000594.

ABSTRACT

Every year approximately two million people in America are affected by healthcare-associated infections (HAIs), with nearly 75,000 of those infections resulting in death. The treatment of HAIs adds nearly \$500 million to US healthcare costs per year. Despite good hand hygiene being the most effective preventive measure, compliance among HCWs still remains low. While there are possible alternative interventions to reduce HAIs—e.g. alternative patient room assignment strategies, changes to hospital infrastructure, vaccination mandates, increased personal protective equipment—implementing them is disruptive and expensive. Therefore our research instead aims to provide a simulation environment that will allow researchers to compare the effectiveness of such interventions in simulation before implementation in the clinical setting.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
INTRODUCTION	1
MATERIALS	2
Dataset Description	2
Limitations	3
METHODS	4
Statistical Distributions	5
Gaussian Mixture Model	5
Zero-Inflated Poisson	7
Simulation	8
Parameter Validation	9
Simulation Design	10
Transmission Model	11
Synthetic Day Generation	12
Replayed Schedules	13
RESULTS & DISCUSSION	13
REFERENCES	17
APPENDIX	19

LIST OF TABLES

Table 1. Table describing healthcare facility groupings	2
Table 2. Comparison Between Facilities 19 and 21	9
Table 3. SQL Database Table Names with Attribute Descriptions	19

LIST OF FIGURES

Figure 1.	Partial Reconstruction of Facility 43 Floor 1, labeled polygons correspond to rooms	3
Figure 2.	Partial Reconstruction of Facility 39 Floor 1	4
Figure 3.	Comparison of fit between Single Normal Density and Gaussian Mixture Model Density	6
Figure 4.	Example code snippet of duration distribution estimation	6
Figure 5.	Example fit of Zero-Inflated Poisson for Nurse Visits	8
Figure 6.	Illustration of Simulation Engine	10
Figure 7.	Estimated Viral Shedding Curve	12
Figure 8.	Graph of infection curves in ICUs using synthetic schedules	14
Figure 9.	Infection incidence over 28-day simulations	14
Figure 10.	Boxplots showing the difference in infection incidence between the consecutive and synthetic schedules	15

INTRODUCTION

Every year approximately two million people in America are affected by healthcare-associated infections (HAIs), with nearly 75,000 of those infections resulting in death. The treatment of HAIs adds nearly \$500 million to US healthcare costs per year [1]. Reducing the spread of HAIs is therefore an area of high interest, with nearly 40% of HAI being attributed to poor hand hygiene on the part of healthcare workers (HCWs) [2], making hand hygiene one of the most effective methods of containment [3]. Most healthcare facilities measure hand hygiene compliance manually, using human observers, though others have adopted automated compliance tracking solutions of various types, such as the system used for data collection in this paper. Despite good hand hygiene being the most effective preventive measure, compliance among HCWs still remains low [4]. While there remain interventions other than hand hygiene—e.g. alternative patient room assignment strategies, changes to hospital infrastructure, vaccination mandates, increased personal protective equipment—that may lead to further protection against the spread of infection, trying to assess the effectiveness of these interventions through direct implementation is costly, time consuming, and unethical given the concern for patient safety. Our research instead aims to provide a simulation environment that allows researchers to evaluate the effectiveness of such interventions in simulation before implementation in the clinical setting.

Agent-based models (ABMs) [5] are an obvious choice for the simulation of disease spread in healthcare settings, but these models rely heavily upon the underlying mathematical models of agent (HCW) movement and disease transmission. As such, existing ABM-driven discrete event simulation studies are often narrow in focus (e.g. examining an individual emergency department)[6][7]. These studies are therefore limited in their ability to be applied in a broader context. To address this limitation, our research aims to provide probabilistic models for healthcare worker behavior across a variety of healthcare settings, derived from empirical data in order to inform more sophisticated simulation efforts.

MATERIALS

Dataset Description

Our work is based on data collected from an automated hand hygiene system consisting of wearable tags with unique badge ID numbers and corresponding instrumented alcohol rub/soap dispensers at known locations within a healthcare facility. Our data set is unique in that it contains data from 25 different healthcare facilities from seven different facility categories as characterized by number of beds, and setting. (Table 1)¹. The data were acquired through a bilateral agreement

Facility Type ID	Facility Type
0	Test
1	Clinic
2	Community (< 200 beds)
3	Teaching (small, < 200 beds)
4	Teaching (medium, 200-299 beds)
5	Teaching (large, >300 beds)
6	Teaching (for profit, 200-299 beds)

Table 1. Table describing healthcare facility groupings

with a privately held commercial entity in the form of three spreadsheets. The first contains a denormalized SQL table of 44 million observations. Each observation represents a single record of the entrance and exit of a specific HCW in a specific location at a specific instance in time. These records also include information about the HCW and their hand hygiene behavior. The second contains approximately 11,000 entries, with each entry describing a single soap/alcohol rub dispenser's location within a healthcare facility. Here, each entry is a single room with a unique room ID Finally, the third file contains approximately 200 records describing units (groupings of rooms in physical proximity and under common administration) and departments (human resource designations). A full list of normalized database table names and descriptions can be found in the Appendix.

¹Other groupings are possible by grouping along other characteristics given in the data such as region, setting, or environment.

Limitations

Although originally designed to monitor hand hygiene compliance, we obtained the dataset to support a much broader and more substantive analysis of HCW behavior. Still, the data have limitations; implementation of the system was not the same across all facilities leading to inconsistent recording of some attributes. For instance, many, but not all, rooms have associated two-dimensional coordinate data, allowing us to partially reconstruct floor plans [8]. However, even within facilities, this information is only available for instrumented rooms, and not for, e.g. hallways, atriums, or other uninstrumented spaces, often leading to floor plans with missing rooms where the geometry can be inferred but is not explicitly included in the dataset. (Fig 1, 2)

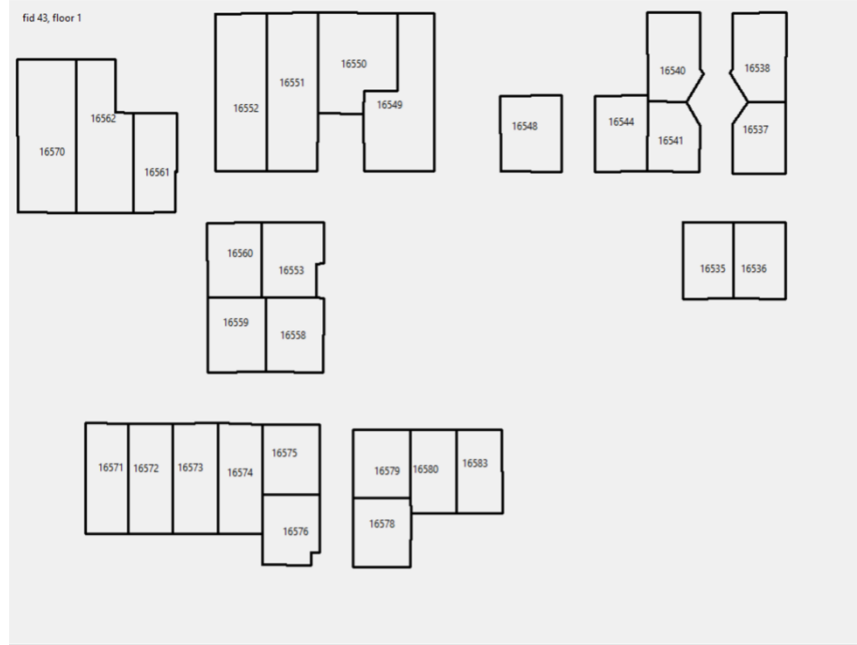


Figure 1. Partial Reconstruction of Facility 43 Floor 1, labeled polygons correspond to rooms

There are other data quality issues, for example, the designation of a room as a patient room is inconsistent. Some facilities set this field to True for all rooms, although it is unclear if this is due to sensors actually being installed only in patient rooms. Other facilities set this field to False for all rooms, possibly indicating the field was not properly initialized. Due to this particular semantic issue, we have elected to ignore this field in our analysis.

In the second file, in which location data is given, we encountered ambiguity in the naming

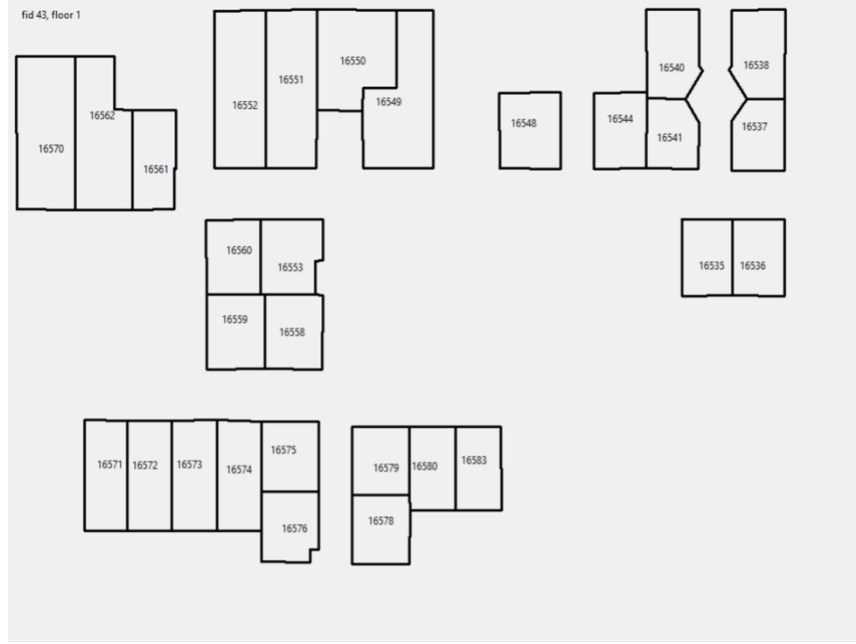


Figure 2. Partial Reconstruction of Facility 39 Floor 1

schema of some units and departments. For example, there were duplicate entries in which a specific unit number was mapped to several different associated names. A similar pattern emerged when investigating the relationship between department ID numbers and their associated names. In the case of duplicates, information was collapsed such that each unit ID is mapped to a single unit name and each department ID is mapped to a single department name. The mappings in units and departments may have some geometric interpretation relating to where units and departments are physically located in a particular facility, but without first-hand knowledge of each facility, it is hard to know what these actually mean.

METHODS

Creating a flexible synthetic hospital will require accurate models of all aspects of HCW behavior. Here, we develop meaningful and generalizable statistical models of two primary characteristics: how often and how long a HCW is visiting an instrumented room.

Statistical Distributions

Gaussian Mixture Model

We begin by exploring the visit duration data. A limitation of the data was that the sensors only recorded a visit if it was at least 60 seconds long, meaning all duration distributions should only be used to model visits of this length or longer. I also found that the data was often heavily tailed, so I transformed the data using natural log. All subsequent distributions are based on the natural log of duration. Although I initially fit the data to power-law distributions, they proved inflexible and fit some job types rather poorly which became problematic when trying to automate the parameter fitting process across the entire dataset. Instead, I altered my approach to try to reflect what I believe to be the underlying mechanism of HCW behavior. My primary assumption is that each HCW jobtype performs a particular set of tasks. For example, a nurse might regularly insert IVs, check vitals, etc. I then assumed that each of these tasks' duration is normally distributed.

n = number of distinct tasks performed during a shift

π_t = proportion of overall tasks made up by a task

μ_t = average duration in seconds of a task

σ_t = standard deviation in seconds of a task

Finally, I assume that each of these tasks make up a fixed proportion of an individual HCW's overall work. Combining these assumptions led to the use of Gaussian mixture models for fitting visit duration density as shown in equation 1.

$$f(t) = \sum_t^n \pi_t * N(\mu_t, \sigma_t^2) \quad (1)$$

Gaussian mixture models are advantageous in that they are more flexible than power-law distributions, they provide a better fit than a single normal distribution, and they accurately represent the distinct set of tasks model of behavior. All parameters provided in this paper were estimated in R using the mclust package which relies on maximum likelihood estimation [9].

Nursing HCWs in Clinic Setting Fit Comparison

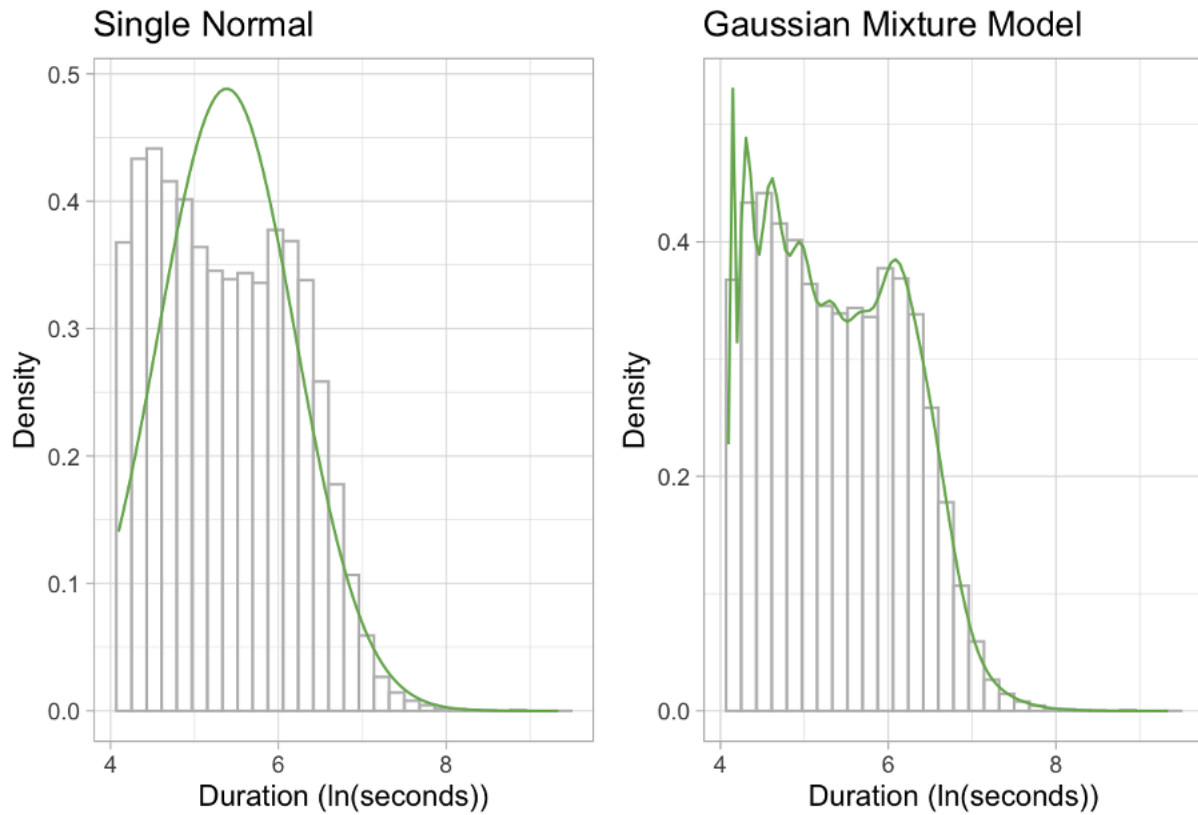


Figure 3. Comparison of fit between Single Normal Density and Gaussian Mixture Model Density

```
hwcOfInterest = dplyr::filter(g, jtid == hcw, daytype == inputs[2])
if(nrow(hwcOfInterest) != 0){
  duration_data = log(hwcOfInterest$duration)
  maxG = min(15, (nrow(duration_data)/20))
  fit_k = Mclust(duration_data, G = 1:maxG)
```

g : facility type

maxG : maximum number of task groups to estimate

Figure 4. Example code snippet of duration distribution estimation

The Mclust function then uses Bayesian inference criteria to pick the correct number of task groups, and maximum likelihood estimation to estimate the mean and variance parameters for each task group. However, I imposed the restriction on the maximum number of groups that could be estimated to prevent overfitting. During testing, I found that most durations were being fit with at maximum 15 distributions, which gave a reasonable upper bound for the maximum distributions as in the testing phase, no facility type/ job type combination was fit with more than 17 task groups. For the lower bound, I imposed the restriction that each normal distribution should be fit using “enough” data, for example fitting a normal distribution with only 1 data point is unreasonable. Although enough is a bit imprecise, it was agreed upon in order to fit a group, at least 20 data points should be used [10]. Thus, leading to the restriction

$$maxG = \min(15, \text{nrow}(\text{duration_data})/20) \quad (2)$$

Results for all facility groups, daytype (weekday/weekend), and HCW jobtype are available on my github [here](#).

Zero-inflated Poisson

To estimate the distribution of visit frequency, I first grouped each observation into the appropriate hour-long bucket. The choice for 1-hour buckets, rather than 30-minute buckets, or some other length of time, stems from considering how we wanted to simulate HCW schedules. We know that in real life hospitals staff differently during the day and night shifts, but with no information about how each hospital is staffed, it was easiest to model number of visits by hour, such that we can catch the switch from day shift to night shift. Next, I created a table with counts where each row represents a different HCW and each column represents an hour-long bucket. While this was helpful in determining how many rooms a particular HCW visited in an hour, I quickly came to the realization that the table had an over-representation of zero values.

The challenge then became determining if a zero was observed due to a HCW not visiting

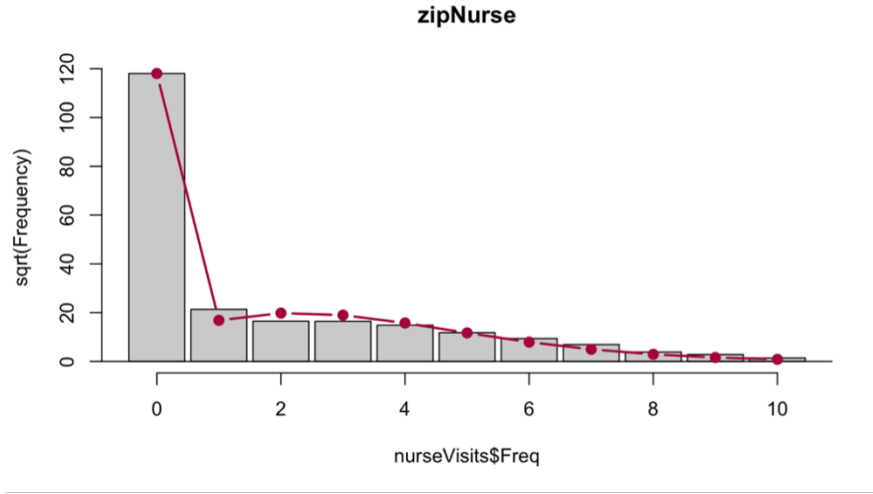


Figure 5. Example fit of Zero-Inflated Poisson for Nurse Visits

any rooms with sensors, a “true zero”—perhaps because they had administrative work to do—or because they were not at the hospital. In order to circumvent this problem, I estimate the visit frequency using a zero-inflated Poisson using the R package `pscl`. This distribution is a mixture of two components, a point mass at zero, and a Poisson distribution measuring counts. Meaning that a zero could come from the point mass or the Poisson count component making it a “true zero” [11]. (Fig 4). The data is then used to estimate parameters from the two component mixture model. For the purposes of simulation, one would want to only use the parameters from the Poisson distribution of the output. Results for all facility groups, daytype (day/night) and HCW job type are available here.

Simulation

We want to show why understanding healthcare behavior reaps actionable insights from simulations. For example do different environments which engender different HCW behaviors also require different interventions? Here we imagine a simple experiment to help validate the usefulness of fine grained HCW behavior models like the ones generated in this paper.

Parameter Validation

On the most basic level we want to compare the difference infection level in two similar units types that come from different facility types. To this end, we choose to analyze the differences between two ICU units in facilities 21 and 19, which are part of facility types 2 and 4, respectively, we choose these two facilities because they are similar on a number of parameters as well as had a similar number of observations in the data set.

Facility	Number of Beds	Environment	Control	Setting	Region	Staff on Duty per Day
21	50-99	Urban	Non-profit	Non-Teaching	West	13
19	200-299	Urban	Non-profit	Teaching	West	52

Table 2. Comparison Between Facilities 19 and 21

Our main area of interest was in comparing the development of the infection outbreak that takes place in the span of 28 days. The pathogen that we chose to simulate is based on an influenza-type respiratory virus. For our initial experiment, I chose to do 10 replicates for each facility, though future experiments could easily use more replicates to better estimate the true underlying behavior. An infection was then simulated in the two different facilities, using two different schedules types, allowing us to compare not only the infection behavior between the two facilities, but the behavior of the simulation between real and synthetic data. Since the staffing levels are so different between each facility, rather than recording the raw number of infections, we focused on the proportion of staff that have been infected. The proportion of infected healthcare workers is then recorded over the 28-day period in order to yield an infection curve.

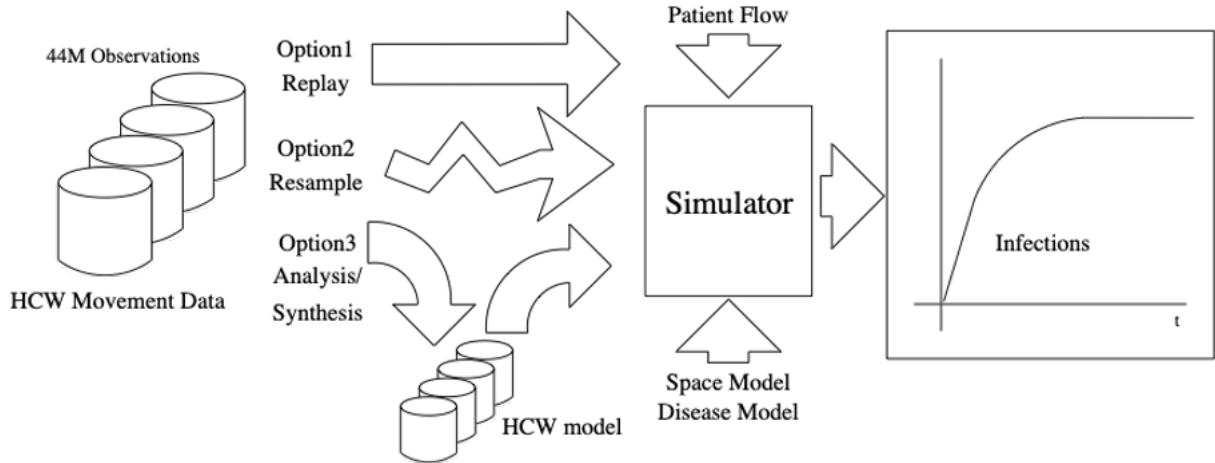


Figure 6. Illustration of Simulation Engine

Simulation Design

The simulation used to produce the results is based on a susceptible-exposed-infected-recovered (SEIR) model, such that patients are no longer susceptible to infection after recovering from a previous infection. For our simulation, the model's infective agents have a 10-day cycle. People are exposed for E days (the incubation period), then infected for I additional days (the symptomatic period). Infectious individuals are those in either the E or I phases. Carrat et al. indicate that E is approximately 2 days and I is approximately 7 days for influenza [12]. Those who are newly infected become infectious at the end of the day that they acquired the infection. The simulation assumes the input of a schedule representing HCW visits to patient rooms, which are assumed to always be filled. The always filled rooms with implemented in the following way; patients are considered for discharge daily with a probability of $1/LOS$ where LOS is length of stay in days. This is true for all patients, except for those infected, who are not eligible for discharge until the infection has run its course, at which point they resume being considered for discharge. If a patient is discharged, a new patient takes their place and is assumed to be uninfected. Infection is possible when infectious and susceptible agents inhabit the same room at the same time. Infection transmission is then simulated based on two factors: the shed rate—as estimated by the number of days since infection of the infectious agent—and the time spent in the room together.

Transmission Model

The probability of infection is subsequently modeled by the following equation:

$$r_i = 1 - \exp(-\mu * \omega * t) \quad (3)$$

Where:

r_i : probability of infection

μ : calibration hyperparameter

ω : shedding rate as estimated by day in infection cycle

t : time in minutes

Our model is also somewhat rudimentary in that it assumes every person is equally susceptible, and there have been no efforts made to decrease the likelihood of infection; therefore, we assume that given enough time spent together, the probability of infection between a susceptible and an infected agent exponentially approaches one hundred percent. Our second assumption is related to the length of time it takes before a person is infected. The hyperparameter informs the model about how infectious the current pathogen is likely to be, as calibrated through probability of infection in a given interval of time. Our experiments were conducted with $\mu = 0.2$, which correspond to $r_i = 0.99$ after 23 minutes, when the infectious agent is at peak viral shedding ($\omega = 1$).

Although viral load is intrinsically linked to the individual infected person's immune response, we assume that viral load can be estimated by the day of infection cycle. As suggested by [12] the viral load typically peaks around the emergence of symptoms and then slowly tapers off. We therefore choose to emulate the shedding curve using the 10-day infection cycle with an assumed onset of symptoms after 2 days. Although the shedding curve (change in viral load) is continuous in nature, we choose to use day of infection cycle—a metric already tracked within our simulation—to set discrete values of ω (Fig. 6) reflective of the shape of the shedding curve as

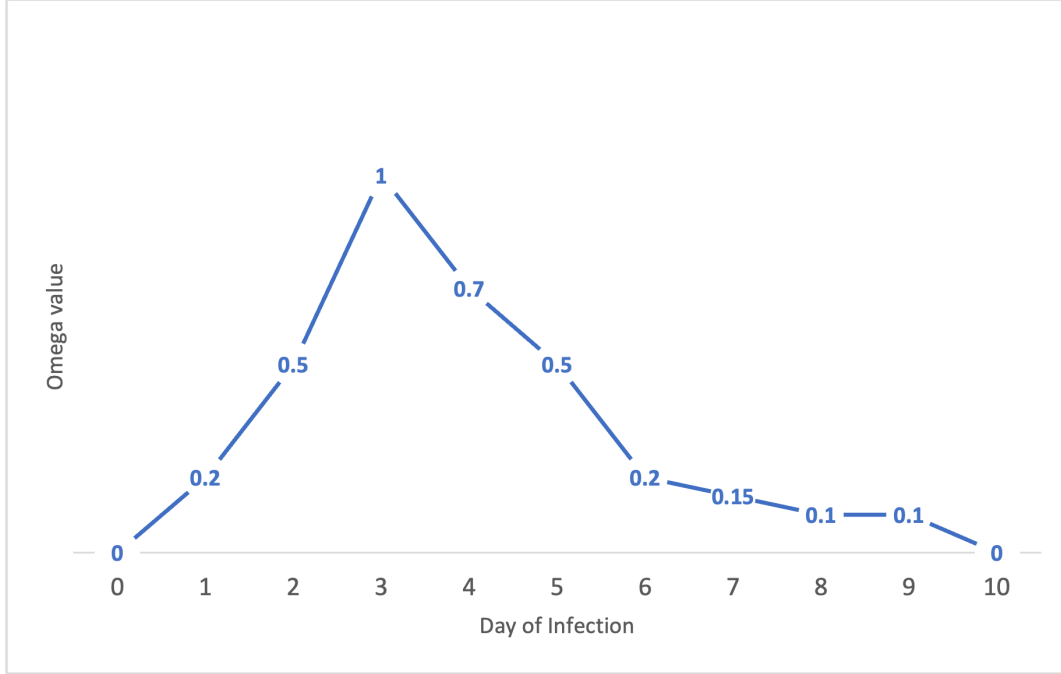


Figure 7. Estimated Viral Shedding Curve

presented in [12].

Synthetic Day Generation

The `syntheticday.py` file forms the backbone of the simulations, as it creates a synthetic schedule of events for a hypothetical day in a hospital. In the schedule, healthcare workers visit rooms in a random order and stay in each room for an amount of time determined by the distributions extracted from the data set. The staffing levels for each facility are based on the real observations of staffing levels on a given day in that facility. The process for generating schedules is iterative in that each individual healthcare worker's schedule is generated hour by hour and then appended to the overall schedule before continuing the generation of the next healthcare worker's schedule. The synthetic schedules simulate each day as a 12 hour shift, with no distinction made between weekends and weekdays. Due to the simplicity of the schedule generation, the schedule must then be modified for use in multiday infection simulations. Since the same HCW identification numbers (HIDs) are reused each day, in order to avoid overworking HCWs and artificially lowering the number of infections recorded in simulation, we developed a system to remap HIDs

wherein we assume that each healthcare worker only works three 12-hour shifts per week. The difficulty with generating synthetic schedules, is that each day uses the same HIDs, meaning every HCW is working all day everyday, therefore the HIDs are overwritten in a greedy fashion to reflect normal weekly working schedules.

Replayed Schedules

As we elected to compare real and synthetic data, it is worth discussing how the real data was chosen. The schedules were generated, by grabbing all data in a particular time 28-day time frame. The same time frame was used for both of the two facilities. As real data is being used, it can be assumed that normal labor practices are respected, and therefore the HIDs are not being overwritten.

RESULTS & DISCUSSION

To directly compare the spread of the infection between the two facilities and to account for different staffing levels between them, I measured the percentage of HCWs infected each day, which when graphed, yield the infection curve for a given simulation replicate. In an effort to reduce the noise produced by generating a single schedule, ten synthetic schedules were generated for each facility, which were then each simulated ten times. *Figure 7* shows the infection curves using the above metric averaged over many simulation runs. Between the two curves, there is a statistically significant difference in the infection behavior from days five through nine, with facility 21—a small community hospital—showing a much slower infection curve but overall higher rate of infection than facility 19—a medium sized teaching hospital.

Since ABMs and discrete event simulation (DES) rely on the accuracy of their underlying mathematical models, it is important not only to provide the parameters of the statistical distributions, but also to make sure that they reflect the dominant trends found in the empirical data. *Figure 8* shows the infection incidence using synthetic schedules as compared to real schedules collected from dataset from the same facilities and units, where each schedule was simulated 10 times.

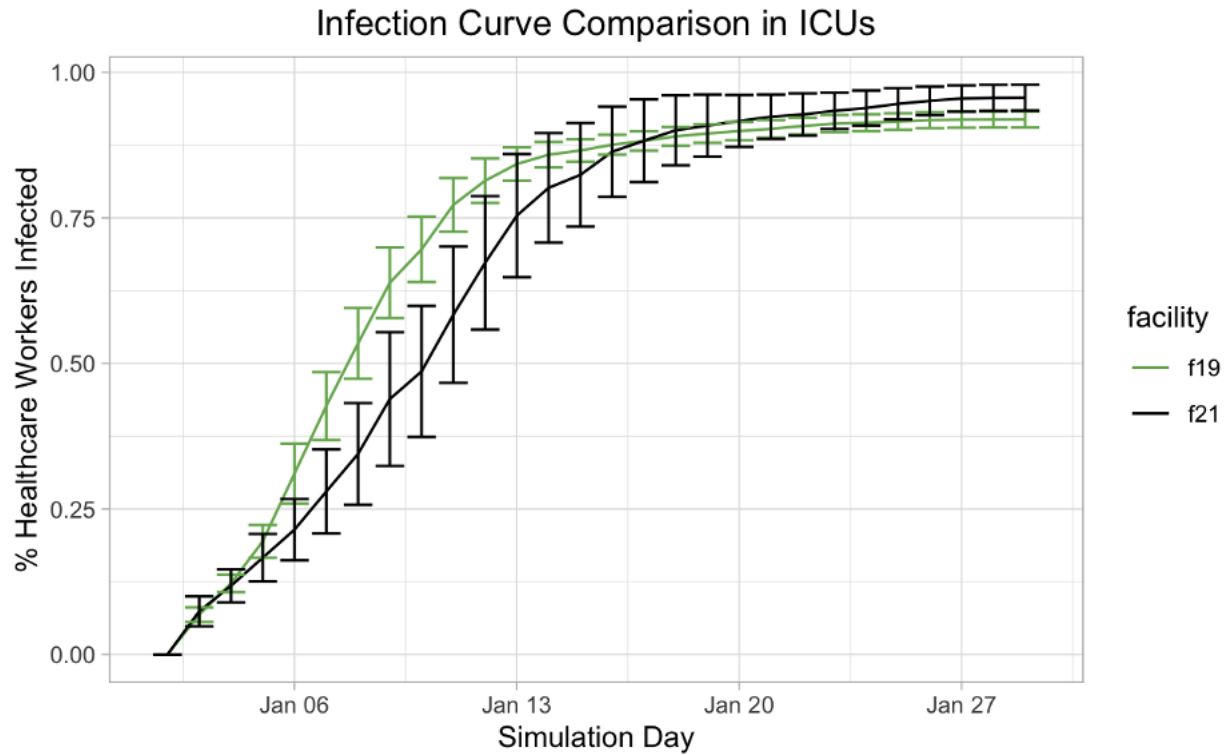


Figure 8. Graph of infection curves in ICUs using synthetic schedules

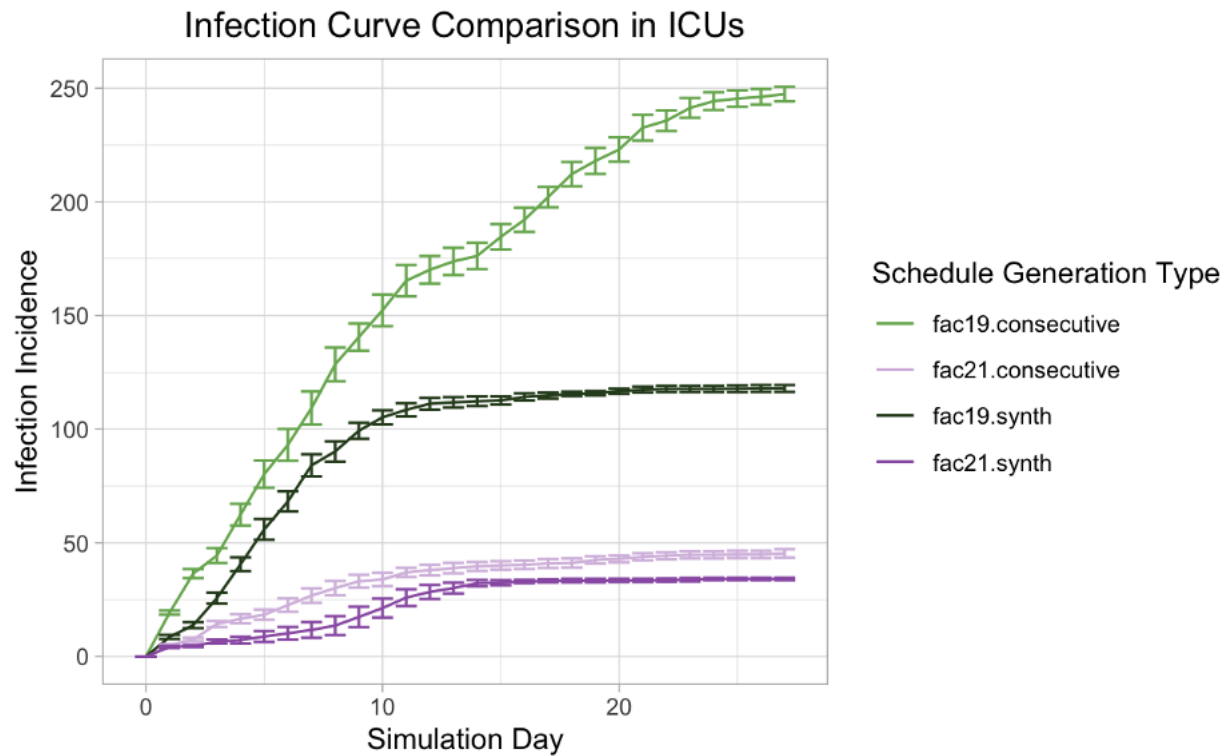


Figure 9. Infection incidence over 28-day simulations

Difference In Infection Incidence

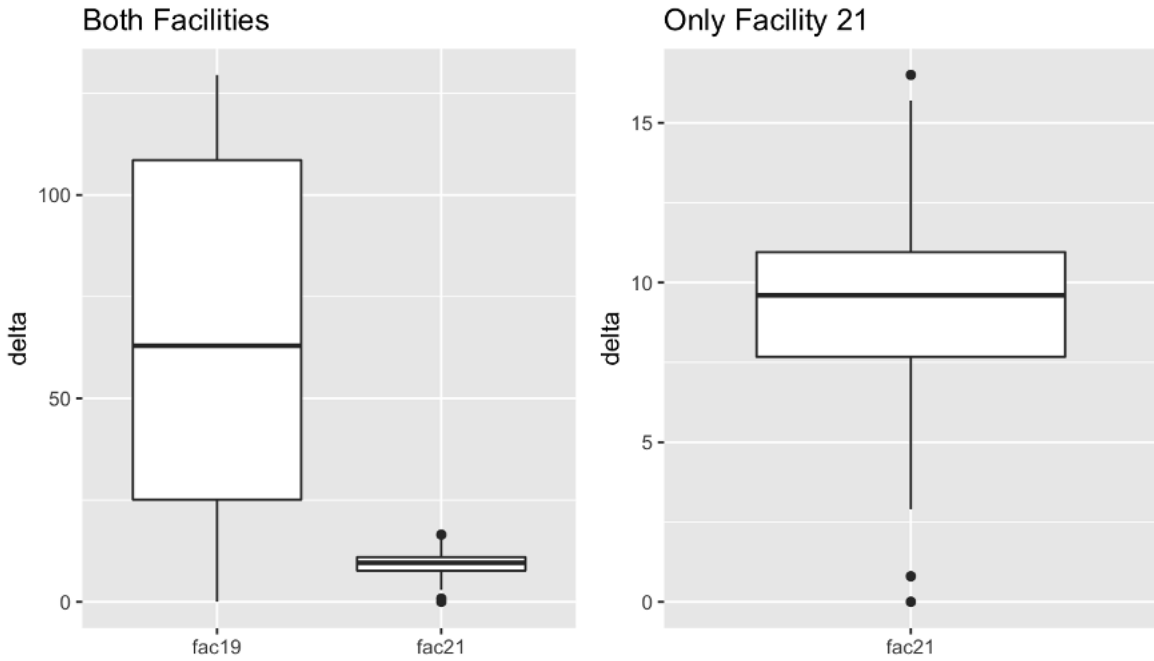


Figure 10. Boxplots showing the difference in infection incidence between the consecutive and synthetic schedules

Although neither facility has overlapping confidence intervals for infection incidence between the synthetically generated and replayed consecutive data, the dominant trends of infection are consistent between simulations run on replayed and synthetic data indicating that when comparing simulations run from synthetic schedules among one another, the differences observed appropriately reflect the differences observed in their real-world counterparts.

While it appears (*Fig. 10*) that the difference in infection incidence between synthetic and consecutive schedules is smaller in smaller facilities, our goal is to minimize the difference between the two schedule types. Future work will include fine tuning the synthetic schedule generation to achieve the previously stated aim. Potential strategies include, analyzing and incorporating the paths HCWs take through hospitals, as currently rooms are visited at random. In addition, incorporating concurrent visits, such that two or more HCWs visit the same room at the same time. Other possible improvements include the incorporation of patient behavior, season, floorplan, and room size into future simulation models. On the transmission model side, the current approach fails

to consider factors such as individual immunity, effective personal protective equipment, and hand hygiene habits which could be topics for future research. There is also potential improvements to be made by tuning the calibration hyperparameter transmission model. During a literature review, we found a study assessing the the risk of airborne transmission of SARS-CoV-2 infection, according to which, when assuming a room of 100m^3 and a natural air exchange rate of half the air in the room per hour, the maximum time until a 99% probability for infection is around 126 minutes [13]. Given our transmission model this would suggest $\mu \approx 0.036$ which is significantly different from the μ used in experimentation with one possible interpretation being that the virus simulated is more infectious than SARS-CoV-2.

Beyond the extension of parameter validation and refinement, future research may entail using the simulation to assess the impact of interventions (such as healthcare worker cohorting, vaccination mandates, increased personal protective equipment, changes to patient room assignments, and changes to hospital infrastructure) on the spread of infections in a variety of hospital settings.

REFERENCES

- [1] R. Assab, N. Nekkab, P. Crepey, P. Astagneau, D. Guillemot, L. Opatowski, L. Temine
”Mathematical models of infection transmission in healthcare settings: Recent advances from the
use of network structured data” *Current Opinion in Infectious Diseases*, vol 30., no. 4, pp.
410-418, 2017
- [2] T. C. Horan, M. Andrus, and M. A. Dudeck “CDC/NHSN surveillance definition of health
care-associated infection and criteria for specific types of infections in the acute care setting”
American Journal of Infection Control, vol. 36, no. 5, pp. 309-332, Jun 2008.
- [3] A. Revelas “Healthcare - associate infections: A public health problem” *Nigerian Medical
Journal*, vol. 53, no. 2, pp. 59-64, 2012.
- [4] J. Randle, M. Clarke, and J. Storr “Hand hygiene compliance in healthcare workers” *Journal
of Hospital Infection*, vol. 64, no. 3, pp. 205-209, 1 Nov 2006.
- [5] R. B. Slayton, J. J. O’Hagan, S. Barnes, S. Rhea, R. Hilscher, M. Rubin, E. Lofgren, B. Singh,
A. Segre, and P. Paul, “Modeling infectious diseases in Healthcare Network (mind-healthcare)
framework for describing and reporting multidrug-resistant organism and healthcare-associated
infections agent-based modeling methods,” *Clinical Infectious Diseases*, 2020.
- [6] M. M. Günal and M. Pidd, “Discrete event simulation for Performance Modelling in Health
Care: A review of the literature,” *Journal of Simulation*, vol. 4, no. 1, pp. 42–51, 2010.
- [7] X. Zhang “Application of discrete event simulation in health care: a systematic review” *BMC
Health Services Research*, vol. 18, no. 1, pp. 1-11, 4 Sept 2018
- [8] Segre, Alberto and Liang Jiazhao ”Geometric Reconstruction of Hospital Rooms from HCW
Locator Data,” (personal communication); Aug 2020.
- [9] Fraley, C. Raftery, A. E. Raftery, L. Scrucca, T. B. Murphy and M. Fop. “mclust: Gaussian
mixture modelling for model-based clustering, classification, and density estimation.” R package
version 5, 2021.
- [10] Sewell, Daniel ”CompEpi Meeting” (personal communication); Jan 2021

- [11] A. Zeileis, C. Kleiber, S. Jackman “Regression Models for Count Data in R.” Journal of Statistical Software, vol. 27, no. 8, pp. 1-25, 29 Jul 2008.
- [12] F. Carrat, E. Vergu, N. M. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, and A.-J. Valleron, “Time lines of infection and disease in human influenza: A review of volunteer challenge studies,” American Journal of Epidemiology, vol. 167, no. 7, pp. 775–785, 2008.
- [13] G. Buonanno, L. Morawska, and L. Stabile, “Quantitative assessment of risk of airborne transmission of SARS-CoV-2 infection: Prospective and retrospective applications” Environmental International, vol. 145, pp. 106112, 1 Dec 2020.

APPENDIX

Table Name	Attributes	Description
depts	did, dname, fid	Information about departments
facilities	fid, size, setting, environment, control, region, system, ftid	Information about the healthcare facility
floors	lid, flevel, fname	Information about physical floor information
ftypes*	ftid, ftype	A table used for grouping similar facilities
hcws	hid, hnum, fid, did, jid	Information about individual healthcare workers and where they work.
jobs	jid, jname, jtid	Links job ids to job names as reported by HR.
jtypes*	jt看id, jname	A table used for grouping similar job names into categories.
map*	mid, uid, fid	Reconstruction of floor layouts
polygons	pid, rid, x, y	Information about dimensions of rooms relative to others on that floor.
rooms	rid, fid, uid, patient	Information about rooms, including patient/non-patient designation.
units	uid, fid, unum, uname, utid	Information about units that rooms belong to.
utypes*	utid, utype	Grouping of units into unit types.
visits	vid, rid, hid, shift, duration, itime, otime, idisp, odisp	Visit information, including hand hygiene.

Table 3. SQL Database Table Names with Attribute Descriptions

** = indicates derived data not provided in original dataset*