RESEARCH ARTICLE   OPEN ACCESS   Check for updates

Stochastic models of influenza outbreaks on a college campus

Subekshya Bidari[a] and Eli E. Goldwyn[b]

[a]Department of Mathematics, University of Colorado Boulder, Boulder, CO, USA; [b]Department of Mathematics, University of Portland, Portland, OR, USA

ABSTRACT

Disease outbreaks on residential college campuses provide an ideal opportunity for mathematical modelling. Unfortunately, publicly available data are rare and many of these outbreaks are relatively small, confounding traditional data-fitting techniques such as least-squares. Using data from three outbreaks during the 2015 and 2017 flu seasons at Trinity College, we fit several SIR-type stochastic models by approximating the likelihood of each model. We find that stochasticity is a key driver in determining the size of the outbreak, and that it strongly depends on the amount of time between the start of the outbreak and the next school holiday. Our results indicate that in order to prevent or limit the size of an outbreak, school closure is likely to be more effective than increasing the vaccination rate. As influenza is a leading cause of negative academic outcomes, these results offer important guidance for school administrators.

1. Introduction

Influenza outbreaks occur regularly and are a significant source of morbidity and mortality. While the prevalence of influenza fluctuates from year to year, in a typical year it is associated with 40,000–50,000 deaths in the USA (Thompson et al., 2003; Viboud, Earn, Simonsen, Dushoff, & Plotkin, 2006). With some notable exceptions (most famously the 1918 H1N1 Spanish flu pandemic), these fatalities predominantly occur in infants, the elderly and immunocompromised individuals. Influenza infection leads to a wide range of symptoms, from serious disorders involving the heart, lungs, brain or muscles to nearly symptomless infection (Zambon, Nicholson, & Wood, 2003). While not usually at high risk of fatality, college students experience substantial morbidity due to influenza infection. Survey data from the American College Health Association regularly lists 'colds/flu/sore throat' as a leading cause of lower exam scores and course grades (American College Health Association, 2016). As campus outbreaks can have high attack rates due to regular exposure in common living quarters, shared restrooms and social activities (National Foundation for Infectious Diseases, 2017; Sobal & Loveland, 1982), college students may be at high risk for infection during a pandemic outbreak.

CONTACT Eli E. Goldwyn goldwyn@up.edu Department of Mathematics, University of Portland, Portland, OR, USA

Supplemental data for this article can be accessed here. https://doi.org/10.1080/23737867.2019.1618744

Fitting models to data is a powerful tool in epidemiology because comparing models and fitting parameters allows for: a better understanding of underlying mechanisms of disease spread (King, Ionides, Pascual, & Bouma, 2008; Koelle & Pascual, 2004), the evaluation of the efficacy of different control methods (Ferguson et al., 2006; Keeling, 2005), understanding of the relative importance of different transmission routes (Atkinson & Wein, 2008) and the importance of social networks on transmission (Eubank et al., 2004; Trapman et al., 2016). Residential colleges and boarding schools provide an excellent opportunity to mathematically model outbreaks due to the relative homogeneity of the population, minimal outside contact and a common health centre (Sobal & Loveland, 1982; Vaidya et al., 2015).

Unfortunately, data on large influenza outbreaks at these types of institutions are rare and often only a relatively small number of individuals are infected (Layde et al., 1980; Pons, Canter, & Dolin, 1980; Schwartz, Morgan, & Lapin, 2015). Fitting mechanistic models to disease outbreak data is difficult even in large outbreaks as these systems are partially observed, nonlinear and stochastic (Ionides, Bretó, & King, 2006). Despite these challenges, there has been success in fitting models using a variety of statistical methods including: least squares prediction (Vaidya et al., 2015), gradient matching (Ellner, Seifu, & Smith, 2002; Hooker, Ellner, De Vargas Roditi, & Earn, 2010), basis expansion (Hooker et al., 2010), iterated filtering (Ionides et al., 2006), synthetic likelihood (Wood, 2010) and data augmented Markov chain Monte Carlo (McKinley, Ross, Deardon, & Cook, 2014). Smaller outbreaks present several additional challenges in applying these techniques, most prominently by magnifying the effects of the stochastic dynamics (process error) and under-reporting (measurement error).

We fit stochastic SIR-type difference equation models to data collected from the Trinity College Health Center (TCHC) from three non-overlapping outbreaks in the Spring 2015 and 2017 flu seasons at Trinity College (Hartford, CT, USA) by simulating these models using MATLAB and approximating the likelihood. Our best-fit models find confidence intervals for the basic reproductive number, R_0, consistent with values generally reported in the literature for seasonal influenza (Biggerstaff, Cauchemez, Reed, Gambhir, & Finelli, 2014). By simulating our best-fit model many times we find that stochasticity leads to a large variance in the total number of infected students and that the time between the introduction of the infection and the next school vacation also plays a large role. Additionally, we examine the effect of outbreak control methods including increasing the vaccination rate and school closure on the size and distribution of the outbreaks.

2. Data

Influenza incidence data were collected by the TCHC. Initially, any student who visited TCHC with symptoms of influenza-like illness (ILI) was tested and confirmed. However, after a run of positive results, any individual with consistent symptoms was considered to be a case of the currently circulating influenza type. The incidence data are anonymized and include date and time of diagnosis. The TCHC is open daily during the school year with the exception of Sundays.

We used data collected by the TCHC from two separate non-overlapping outbreaks during the spring semester of 2015 and one outbreak in the spring semester of 2017. The first

outbreak infected 17 students between 4th February and 20th February 2015 and was confirmed as influenza A (H3N2). The second outbreak infected 25 students between 22nd April and 8th May 2015 and was confirmed as Influenza B. The third outbreak infected 33 students between 3rd February and 17th February 2017 and was also Influenza B. There were no reported cases for several days after each outbreak. The first and third outbreaks ended at the start of Trinity Days (a four day school vacation), while the second ended at the beginning of reading days. This leads us to hypothesize that transmission for each of the outbreaks was interrupted by these scheduled school closures and that the outbreaks did not burn out on their own. The TCHC reports vaccinating 295 students in 2015, but stopped tracking vaccinations after that year. According to the Registrar, there were a total of 2287 students at Trinity College.

Under-reporting of infected cases in disease outbreaks can bias model fits and parameter estimation (Gamado, Streftaris, & Zachary, 2014). In order to estimate the proportion of students who would visit the TCHC (as opposed to another medical provider or none at all) for influenza vaccination or for influenza diagnosis after experiencing ILI, we created an online survey using the website www.surveymonkey.com. The survey consisted of two questions: 'Would you visit the health center on campus if you have flu-type symptoms (headache, fever and so on) and if the services were free?' Answer choices: '(i) Yes (ii) No' and 'Did you get flu shot last year?' Answer choices: '(i) Yes, at the campus health center. (ii) Yes, somewhere else. (iii) No '. We sent the survey to only the upperclassmen (sophomores, juniors and seniors) at Trinity College on September 12, 2016 who number 1722 of the total student population of 2287. We didn't send the survey to the incoming freshman class as they were just arriving on campus and were not yet familiar with the TCHC. The phrase 'if the services were free' was added because between the outbreaks in the Spring of 2015 and sending out the survey, the TCHC began charging a copay for appointments. A total of 225 students responded to the survey: 83% reported that they would visit the TCHC if they experienced ILI symptoms and 52% reported that they received a vaccination (21.8% at the TCHC and 30.2% elsewhere). The survey reported vaccination rate at the TCHC of 21.8% is substantially higher than the 12.9% rate recorded by the TCHC. The most likely explanations are that the population of students responding to the survey are biased towards being more likely to visit the health center for vaccinations, students incorrectly remembered their vaccinations, or (since the survey was sent out after the outbreaks) the vaccination rate increased after the school publicized the Spring 2015 outbreaks. We assume the same over-reporting rate for external vaccinations and therefore reduce the 30.2% rate described in the survey to 17.9%. Adding these two percentages gives a total vaccination rate of 30.8% (704 students). This fits within the 8–39% estimated range for vaccination rates among all college students (National Foundation for Infectious Diseases, 2017). In the model, we assume a reporting rate of illness at the 83% found in the survey.

3. Model

We compare the results from four different nested models describing the influenza outbreaks at Trinity College. Each model follows the SIR compartment paradigm (Kermack & Mckendrick, 1927) with stochastic difference equations and a time step of 1 h. Due to the small population and outbreak size, we restrict the state variables to integer values and

define the probabilities of movement between compartments as binomial random variables. This is the well-known Reed-Frost formulation (Vynnycky & White, 2010). This added stochasticity avoids issues that come from rounding the populations and helps capture the inherent randomness of the disease transmission process. We use the MATLAB function 'binornd' to draw binomial random variables.

Our models include a susceptible class, S, an infectious class, I, and a recovered class, R. The CDC reports that symptoms start 1–4 days after initial infection and that most healthy adults become infectious 1 day before symptoms develop and up to 5–7 days after becoming sick (Centers for Disease Control and Prevention, 2016). To account for these time lags, we add a latent class of infected but not infectious individuals, E, and we subdivide the infectious class into I_A and I_S to distinguish between asymptomatic and symptomatic infectious individuals. We assume that both of these infectious classes equally contribute to transmission, however, individuals only visit the TCHC once symptoms occur. We call our model with all of the compartments listed above SEI_2R (see Supplementary equation (1) for the mathematical formulation):

$$S \xrightarrow{\lambda} E \xrightarrow{\sigma} I_a \xrightarrow{\gamma_S} I_s \xrightarrow{\gamma_R} R$$
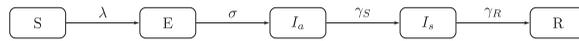
For an individual, each of the above transitions are Bernoulli random variables with the above listed transition probabilities. With the exception of the force of infection, λ, the per capita transition probabilities are constant throughout the outbreak. The average recovery rate of a symptomatic infectious individual is γ_R, the average rate of an infected individual becoming symptomatic is γ_S, and the average rate that a newly infected individual becomes infectious is σ. The expected time spent in each of the above compartments is therefore $1/\gamma_R$, $1/\gamma_S$ and $1/\sigma$ respectively. Since we are primarily concerned with examining the size of the outbreak, the effects of vaccination and school closure, and due to the limited size and length of the outbreaks, we only fit the parameter p, while using the CDC estimates for the other rate parameters. The stochastic force of infection is a function of the number of currently infectious individuals and is calculated in every time step as Vynnycky and White (2010):

$$\lambda = 1 - (1 - p)^I \tag{1}$$

where the parameter p is the probability of effective contact per time step and I is the total number of infectious individuals ($I_a + I_s$). Note that the basic reproductive number is

$$R_0 = \frac{Np}{\gamma} \tag{2}$$

where N is the number of students at Trinity College and $1/\gamma$ is the total time spent in the infectious period. For the models with two I classes $1/\gamma = 1/(\gamma_S + \gamma_R)$.

Theoretical work by Wearing et al. (2005) demonstrated that ignoring the latent period or making the assumption of exponentially distributed latent and infectious periods results in underestimation of the basic reproduction number of the infection from outbreak data. To check the importance of the latent and asymptomatic periods, we compare the fits of four candidate models to the outbreak data: SEI_2R, SI_2R, $SEIR$ and SIR. Each of the latter three models can be constructed by removing classes from the SEI_2R model as indicated by the letters in the model name. The SI_2R model assumes no latent time between when

an individual contracts influenza and when she becomes infectious, therefore the *E* class is removed. The *SEIR* and *SIR* models ignore the asymptomatic infectious period, I_A, assuming that the onset of symptoms coincides with the individual becoming infectious. To test the importance of the distribution of latent and infectious periods, we compare the fit of the models described above with models that include multiple subcompartments in each of the latent and infectious compartments. Adding subcompartments changes the probability distribution of time spent in a compartment from a geometric distribution to a negative binomial distribution (the discrete analogs of exponential and gamma distributions respectively).

3.1. Model assumptions and initial conditions

Trinity College has a student population of 2287; we assume a constant population as students rarely leave during the semester. In the first outbreak, two students were diagnosed by the TCHC on the first day, followed by one student the next day. During the second outbreak, two students were diagnosed on the first day, followed by two students on the next day. The third outbreak saw one student diagnosed on the first day followed by three students on day 3. We use the students visiting the TCHC on the first day as the initial number of symptomatic infectious students, and the subsequent student visits mentioned above as the initial asymptomatic and latent students as appropriate. The CDC published vaccine efficacy for the H3N2 strain (February 2015 outbreak), for Influenza B (April 2015 outbreak and 2017 outbreak), as 23%, 45% and 73%, respectively (Centers for Disease Control and Prevention, 2016, 2018). As described in the previous section, based on data from the TCHC and survey responses, we assume that 704 were vaccinated. To take these successful vaccination probabilities into account, we assume the initial number of individuals in the removed class is a binomial random variable with the above probabilities and 704 trials.

 Using the above described CDC estimates for recovery and latent periods for seasonal influenza, we assume a mean 5 day period in the infectious periods for all models and let the latent and asymptomatic periods to both be 1 day. Note that for the SEI_2R and the SI_2R models this yields an average of 4 days in the symptomatic infectious stage and 1 day in the asymptomatic infectious stage. As faculty and staff do not live with the students and have fewer close interactions, we do not include them in our model (Vaidya et al., 2015). Additionally, while influenza has multiple transmission pathways, we assume transmission occurs only through direct interaction with an infected individual and we ignore transmission through fomites (Li, Eisenberg, Spicknall, & Koopman, 2009).

3.2. Parameter estimation and model selection

Because the size of our data sets is limited (in terms of length of outbreak and number infected), and because our primary goal is to explore the size and distribution of outbreaks and to examine control methods, we assume that all the rate parameters are constant with the above values and only fit the effective contact rate *p*. Equation 2 shows that since the recovery rate and population size are assumed to be constant, fitting *p* is equivalent to fitting R_0. Stochastic models yield a distribution of results, so we simulate each of our four stochastic models 10,000 times for each submodel, different numbers of subcompartments in the latent and infected classes, and over a range of values for *p*. We fit the model results to newly

symptomatic individuals as the data collected by the TCHC are date and time of diagnosis, as opposed to total number of infectious students. To account for under-reporting, we use the survey data which indicates that only 83% of symptomatic students would visit the TCHC. For simplicity, we ignore the possibilities of false positive diagnoses and additional under-reporting coming from asymptomatic individuals.

The data do not include the timing of when individuals are infected or when they recover, they only include the timing newly infected individuals, and are also likely missing unreported cases. Because of these factors, calculating the likelihood function for a given model and parameter value is difficult (McKinley et al., 2014). Additionally, over the course of the outbreak, multiple days have only zero or one newly infected individual; this leads to least squares methods producing large residuals when a model occasionally produces a very large outbreak, while simulations with smaller outbreaks or outbreaks that prematurely burn out have residuals bounded by the data. Therefore, finding the least squares error biases the model to smaller outbreaks and dramatically underestimates R_0 yielding values of 0.78, 0.95 and 1.14 for the each of the three outbreaks respectively, leading to a high probability of rapid die-out of the disease (see Supplementary Table 5).

We use Approximate Bayesian Computation (ABC) methods to approximate the likelihood of a given model and R_0 value by calculating the proportion of simulations that exactly predict the data on each day and then taking the product of those results. Since the number of new cases on each day is relatively small we exactly match the data and simulated results as opposed to creating a new metric to allowing for some distance between the two (McKinley et al., 2014). Note that while we use ABC to approximate the likelihood, because of the limited data we choose not to take a Bayesian Approach to parameter fitting.

We use a negative binomial model to account for under-reporting (Fraser et al., 2009). The data are treated as one draw of a binomial distribution with the number of trials (actual number of newly infected students) as unknown, the number of success as the number of reported cases, and the probability of success as the reporting rate. The likelihood is now approximated by the probability that the simulation result was the same as the actual and unknown number of newly infected students each day.

4. Results

We approximate the log-likelihood for each of the three outbreaks and plot it as a function of R_0 in Figure 1. Because there is no reason to assume that each outbreak would have the same distribution of infectious period, the log-likelihood is maximized over different numbers of subcompartments for each of the four different models. Supplementary Tables 1–3 summarize the best-fit results for all models, R_0 values, and differing number of subcompartments. The maximum log-likelihood and its corresponding R_0 are summarized in Table 1. The values in the table are the maximum value in the highest curve in Figure 1, R_0 is 1.26, 1.50 and 2.06 for each of the three outbreaks respectively.

In order to find confidence intervals for the R_0 for each outbreak, we use the likelihood ratio test. Since we are fitting one parameter, the 95% confidence interval is all log-likelihoods within 1.96 points of the maximum log-likelihood. Based on 10,000 simulations, the 95% confidence intervals for R_0 are $(1.10, 1.82)$, $(1.39, 2.37)$ and $(1.6, 2.97)$, for each of the three outbreaks. Recent estimates of R_0 for a variety of seasonal influenza
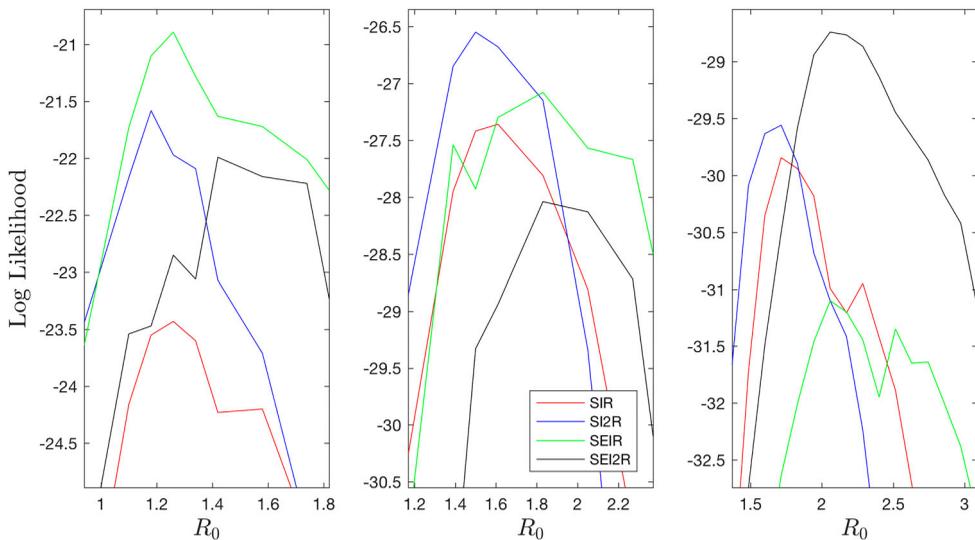
Figure 1. The log-likelihood for each of the four models maximized over various numbers of subcompartments. Results for the first outbreak (left), second outbreak (centre) and third outbreak (right).

Table 1. Maximum-likelihood and corresponding R_0 values for the three outbreaks. The best fit for outbreak 1 was with the *SEIR* model using 10 subcompartments, outbreak 2 with the SI_2R model and 1 subcompartment, and outbreak 3 with the SEI_2R and 4 subcompartments.

Outbreak 1		Outbreak 2		Outbreak 3	
R_0	Max Lhood	R_0	Max Lhood	R_0	Max Lhood
1.26	−20.89	1.50	−26.55	2.06	−28.74

strains range between 0.9 and 2.1 (Coburn, Wagner, & Blower, 2009), consistent with our best fit values and the above confidence intervals.

Using these confidence intervals and comparing the four models, we find that in the first outbreak, while the maximum likelihood fit occurs with the *SEIR* model, it is not significantly better than the SI_2R and SEI_2R models. The *SIR* model, however, offers a significantly worse fit. In the second outbreak, the SI_2R model has the maximum likelihood, but is not significantly better than any of the other three models. In the third outbreak, the SEI_2R model performs significantly better than the *SEIR* model, but not the other two. The number of subcompartments (i.e. the shape parameter on the distribution of time spent in the latent and infectious classes) does not significantly affect the fit of the model for any of the three outbreaks.

To visually examine the fit, we plot time series results from 10,000 simulations using the best-fit model, R_0, and number of subcompartments for each outbreak in Figure 2. For clarity, we only display the 5th, 25th, 50th, 75th and 95th percentile trajectories as measured by total number of infected students. Note that the percentiles are calculated in terms of the total number of infected students throughout the outbreak, therefore it is possible for a lower percentile trajectory to have more infected students than a higher percentile trajectory on a given day. In general, the simulations capture much of the qualitative behaviour of
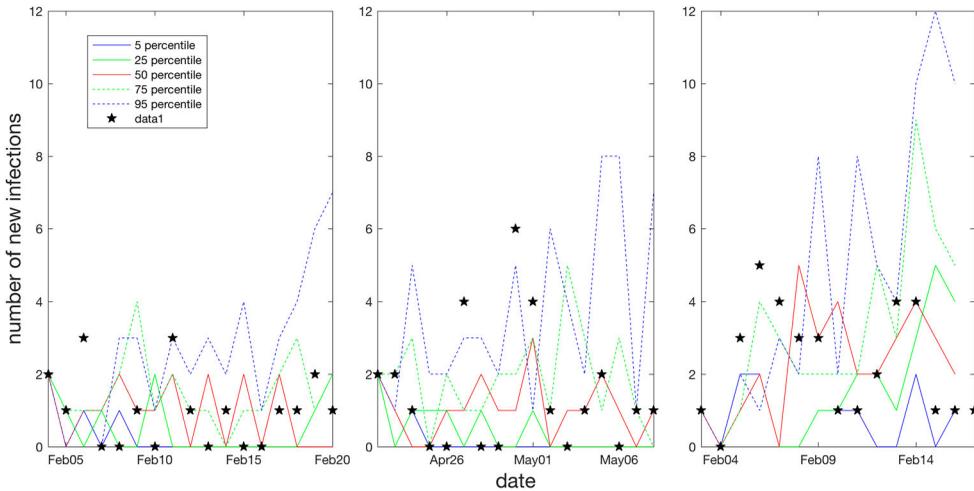
Figure 2. Plot showing 5 percentile (solid blue), 25 percentile (solid green), 50 percentile (red), 75 percentile (dashed green) and 95 percentile (dashed blue) of total infected using 10,000 simulation and $R_0 = 1.26$ for the H3N2 strain (left), $R_0 = 1.50$ for Influenza B (centre) and $R_0 = 2.06$ for third outbreak (right). The black stars represent the TCHC data.

each outbreak, though they occasionally struggle to capture 1-day spikes from 0 to several new cases early on in the process, and sometimes overestimate the number of cases late in the outbreak. This could be due to a change in behaviour of the students after receiving a warning email from the TCHC that would essentially lower p. Note that the simulations plot the total number of newly infected students whereas the data is in terms of reported cases.

Using the same 10,000 simulations from our best fitting model and the actual time span of the infection, we plot histograms describing the total number of infected individuals in Figure 3. The median number of infected individuals are 19, 26 and 41 for the three outbreaks with 95% of our simulations in the intervals $(4, 50)$, $(3, 88)$ and $(9, 92)$. The lower bounds of these simulations represent the case where the disease rapidly died out while the upper bounds describe a scenario where nearly 5% of the student body is infected in each of the last two outbreaks. If we define an outbreak as occurring when at least 10 students become infected, then an outbreak fails to occur 23%, 22% and 6% of the time for the three outbreaks, respectively. Recall that the data from the TCHC reported 17, 25 and 33 infected students during each of the three outbreaks. Since the model tracks the total number of newly infected students, while the data presumably includes under-reporting, we would expect it to overapproximate the data.

The Trinity College community was fortunate that, due to planned school breaks, all three influenza outbreaks were interrupted even though there were still infectious students. In order to determine plausible worst-case scenarios for each outbreak, we find that the longest continuous period of time without a major break in classes is 40 days in the fall semester and 50 days in the spring semester. We use our best fit models for that extended 50 days time line and plot a histogram for the total number of infected in Figure 4. The median number of infected individuals for this maximal time frame is 78 for the first outbreak, 199 for the second outbreak and 998 for the third outbreak. 95% of our simulations
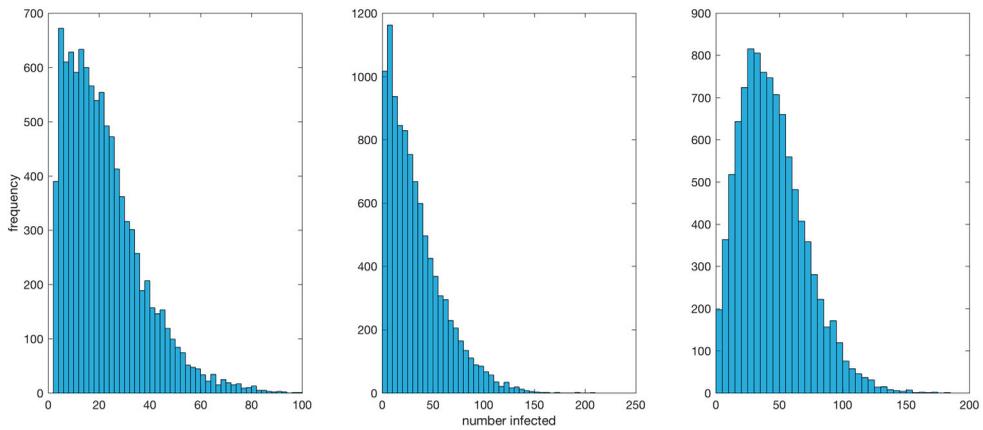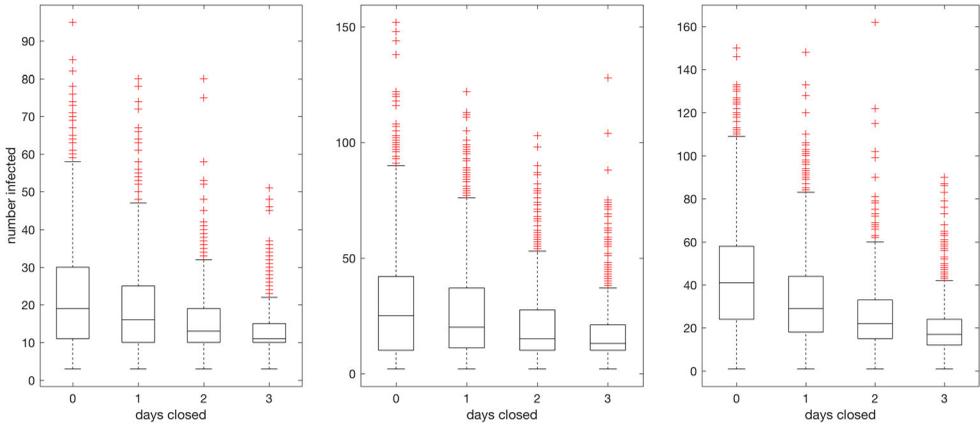
Figure 3. Histograms for the total number of infected students based on 10,000 simulations using our best-fit model and parameters with $R_0 = 1.26$ for the H3N2 strain (left), $R_0 = 1.50$ for Influenza B (centre) and $R_0 = 2.06$ for third outbreak (right). The actual number of students diagnosed by the TCHC is 17 for the H3N2 strain, 25 for Influenza B and 33 for the third outbreak.
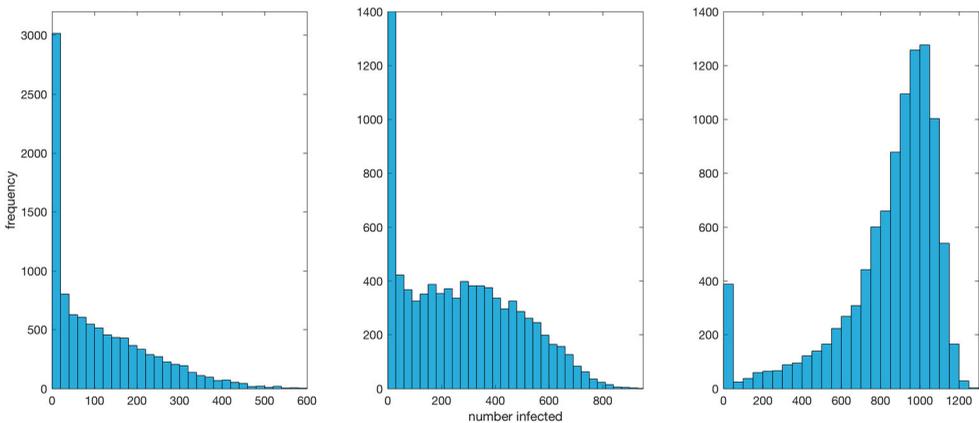

Figure 4. Histograms for the total number of infected students based on 10,000 simulations using the same model and parameters as Figure 3 assuming no school holiday and a maximum 50 day outbreak.

were in the intervals $(4, 347)$, $(3, 632)$ and $(190, 1116)$. Here we see that the medians are roughly 4 to 20 times higher when the outbreak starts just before the longest uninterrupted section of school than in the shorter actual time frame and the upper bounds of the 95% intervals are roughly 7 to 10 times larger.

Based on our best-fit estimates of R_0 and the vaccine efficacy reported by the CDC, we calculate that vaccination rates would have to be 89.7%, 74.1% and 73.9% to reduce the effective R_0 to 1 and therefore prevent each outbreak. These uptake rates are over double the estimated vaccination rate across colleges nationwide. Another strategy to limit the size of an outbreak is to temporarily close the school. The World Health Organization suggests that school closures should be considered in the case of a severe pandemic (Bell et al., 2006). Both empirical reviews (Jackson, Vynnycky, Hawker, Olowokure, & Mangtani, 2013) and

Figure 5. Boxplots from 1000 simulations for the total number of infected students for each of the three outbreaks based on closing the school for 0, 1, 2 or 3 days after 10 students were infected.
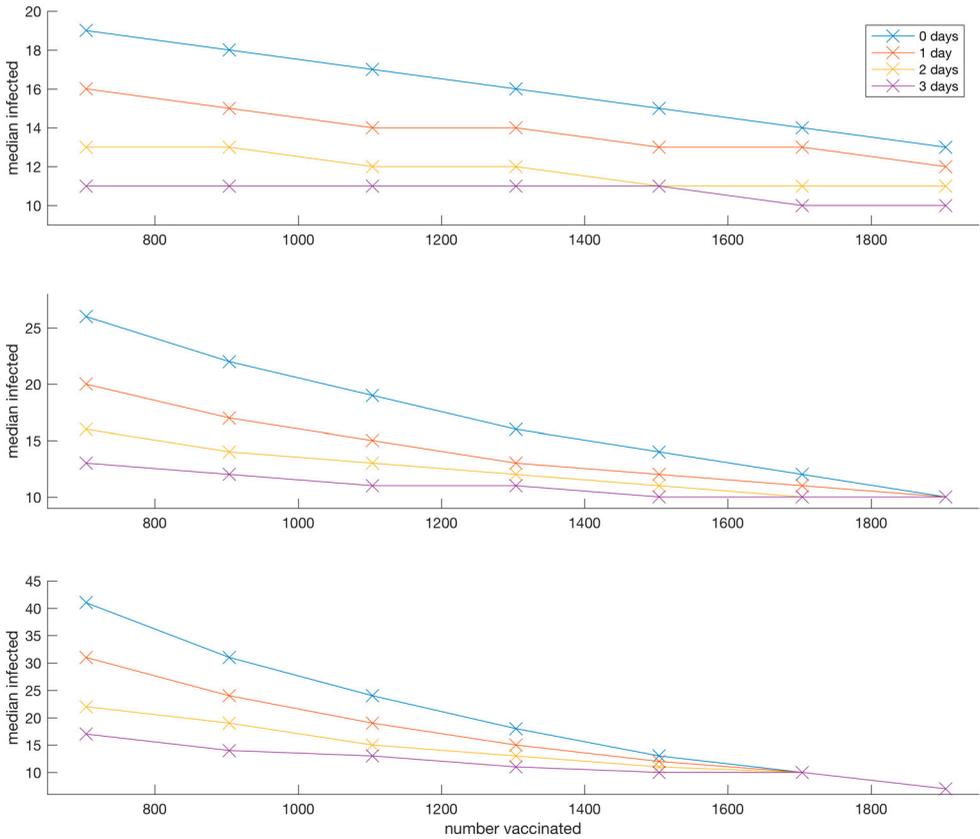
Figure 6. Plot showing the median number of infected students as the total number of students vaccinated is increased. The four different curves describe the results for 0, 1, 2 and 3 days of school closure.

simulation-based studies have studied the effect of school closure on the course of an outbreak (Ferguson et al., 2006). We test the effect of closing Trinity College for up to 3 days by assuming that no transmission happens during that time. Since the TCHC sends out a college-wide email after 10 students are infected, we assume school closure would also occur after the 10th student is infected. The boxplots in Figure 5 demonstrate the decrease in number of students infected for each day of school closure. As one would expect, the median, 75th percentile, standard deviation and maximum number of students infected all decrease every additional day the school is closed. Though even after closing the school for 3 days, there is still a chance for large outbreaks.

We also examined the effect of school closure alongside an increased vaccination rate on the outbreak. Figure 6 shows that the vaccination rate has to be greatly increased to have the same effect as several days of school closure. In the first outbreak, additional 600 students need to be vaccinated for the median to decrease the same amount as it does if the school is closed for a single day and a total of 1200 additional students need to be vaccinated to have the same effect as 2 days of school closure.

5. Discussion

College students are an overlooked population in terms of health implications due to seasonal influenza. While mortality rates remain very low, influenza causes significant morbidity in otherwise healthy college students and leads to worse academic outcomes (American College Health Association, 2016). Higher attack rates at residential colleges may also lead to larger outbreaks in the case of pandemic influenza. Fortunately, in addition to the standard vaccination and pharmaceutical control methods, residential colleges have the ability to limit or prevent large outbreaks by closing the campus (Ciavarella, Fumanelli, Merler, Cattuto, & Ajelli, 2016; Fumanelli, Ajelli, Merler, Ferguson, & Cauchemez, 2016). As improving vaccination rates among college students has repeatedly shown to be very difficult (National Foundation for Infectious Diseases, 2017), and the high vaccination rates necessary to prevent outbreaks may be unrealistic, the closing of campus could be a valuable control method for a large outbreak.

Our best-fit models indicate a wide range of possible outbreak sizes due to the inherent stochasticity in small populations. The incidence data indicate that none of the three outbreaks burned out on their own, but ended due to planned school holidays. Our simulations demonstrate that the size of the outbreak is largely determined by the length of time from the index case until the next school holiday, making the time between the start of an outbreak and the next school vacation crucial in predicting the size of the outbreak. The simulations also indicate that school closure is likely to be a more effective method of control than increasing the vaccination rate. Therefore, in addition to the progression of an outbreak, the academic calendar gives immediate guidance as to whether school administrators should consider school closure as an option.

Modelling disease outbreaks for non-mandatory reportable diseases like influenza in the population at large is a difficult task due to complicated social networks, heterogeneity in individual susceptibility, gathering data from multiple health centres and under-reporting. Outbreaks in total institutions like boarding schools, residential colleges, military bases, nursing homes, and prisons, however, provide excellent opportunities for modelling and data fitting (Sobal & Loveland, 1982). Unfortunately, incidence data on many of these

outbreaks is not publicly available and many of these outbreaks are relatively small, which makes fitting the models to the data difficult. Nonetheless, extending this work by combining it with other college outbreaks for which data exists (Schwartz et al., 2015; Vaidya et al., 2015) and focusing on control methods (Nichol et al., 2010) could provide important guidance for colleges or other institutions trying to control influenza or other outbreaks. Additionally, incorporating more detailed data that include social network information could substantially improve the fit of the model and therefore its predictions and the efficacy of control recommendations (Funk, Salathé, & Jansen, 2010; Salathe & Jones, 2010).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

American College Health Association. National college health assessment. http://www.acha-ncha. org/pubs_rpts.html. Accessed: 2016-10-12.

Atkinson, M. P., & Wein, L. M. (2008). Quantifying the routes of transmission for pandemic influenza. *Bulletin of Mathematical Biology*, *70*(3), 820–867.

Bell, D., Nicoll, A., Fukuda, K., Horby, P., Monto, A., Hayden, F., . . . van Tam, J. (2006). Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerging Infectious Diseases*, *12*, 88–94.

Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., & Finelli, L. (2014). Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: A systematic review of the literature. *BMC Infectious Diseases*, *14*(1), 480.

Centers for Disease Control and Prevention. Key facts about influenza. http://www.cdc.gov/flu/keyfacts.htm. Accessed: 2016-10-12.

Centers for Disease Control and Prevention. 2014–2015 flu season drawing to a close. http://www.cdc.gov/flu/news/2014-2015-flu-season-wrapup.htm. Accessed: 2016-10-12.

Centers for Disease Control and Prevention. CDC reports this season's flu vaccine reducing risk by nearly half. https://www.cdc.gov/flu/spotlights/flu-vaccine-reducing-risk.htm. Accessed: 2018-6-12.

Ciavarella, C., Fumanelli, L., Merler, S., Cattuto, C., & Ajelli, M. (2016). School closure policies at municipality level for mitigating influenza spread: A model-based evaluation. *BMC Infectious Diseases*, *16*(1), 576.

Coburn, B. J., Wagner, B. G., & Blower, S. (2009). Modeling influenza epidemics and pandemics: Insights into the future of swine flu (H1N1). *BMC Medicine*, *7*(1), 30.

Ellner, S. P., Seifu, Y., & Smith, R. H. (2002). Fitting population dynamic models to time-series data by gradient matching. *Ecology*, *83*(8), 2256–2270.

Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*(6988), 180–184.

Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., & Burke, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature*, *442*(7101), 448–452.

Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerk hove, M. D., Hollingsworth, T. D., . . . Jombart, T. (2009). Pandemic potential of a strain of influenza a (H1N1): Early findings. *Science*, *324*(5934), 1557–1561.

Fumanelli, L., Ajelli, M., Merler, S., Ferguson, N. M., & Cauchemez, S. (2016). Model-based comprehensive analysis of school closure policies for mitigating influenza epidemics and pandemics. *PLOS Computational Biology*, *12*(1), 1–15.

Funk, S., Salathé, M., & Jansen, V. A. A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: A review. *Journal of The Royal Society Interface*, *7*(50), 1247–1256.

Gamado, K. M., Streftaris, G., & Zachary, S. (2014, September). Modelling under-reporting in epidemics. *Journal of Mathematical Biology*, *69*(3), 737–765.

Hooker, G., Ellner, S. P., De Vargas Roditi, L., & Earn, D. J. D. (2010). Parameterizing state–space models for infectious disease dynamics by generalized profiling: Measles in ontario. *Journal of The Royal Society Interface*, *8*(60), 961–974.

Ionides, E. L., Bretó, C., & King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *103*(49), 18438–18443.

Jackson, C., Vynnycky, E., Hawker, J., Olowokure, B., & Mangtani, P. (2013). School closures and influenza: Systematic review of epidemiological studies. *BMJ Open*, *3*(2), e002149.

Keeling, M. J. (2005). Models of foot-and-mouth disease. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1569), 1195–1202.

Keeling, M. J., Wearing, H. J., & Rohani, P. (2005). Appropriate models for the management of infectious diseases. *PLoS Medicine*, *2*(7), 00–00.

Kermack, W. O., & Mckendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, *115*(772), 700–721.

King, A. A., Ionides, E. L., Pascual, M., & Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature*, *454*(7206), 877–880.

Koelle, K., & Pascual, M. (2004). Disentangling extrinsic from intrinsic factors in disease dynamics: A nonlinear time series approach with an application to cholera. *The American Naturalist*, *163*(6), 901–913.

Layde, P. M., Engelberg, A. L., Dobbs, H. I., Curtis, A. C., Craven, R. B., P. L. Graitcer, . . . Noble, G. R. (1980). Outbreak of influenza A/USSR/77 at marquette university. *Journal of Infectious Diseases*, *142*(3), 347–352.

Li, S., Eisenberg, J. N. S., Spicknall, I. H., & Koopman, J. S. (2009). Dynamics and control of infections transmitted from person to person through the environment. *American Journal of Epidemiology*, *170*(2), 257–265.

McKinley, T. J., Ross, J. V., Deardon, R., & Cook, A. R. (2014). Simulation-based Bayesian inference for epidemic models. *Computational Statistics & Data Analysis*, *71*, 434–447.

National Foundation for Infectious Diseases. Addressing the challenges of influenza vaccination on us college campuses. http://www.nfid.org/publications/reports/college-flu-summit-report.pdf. Accessed: 2017-1-12.

Nichol, K. L., Tummers, K., Hoyer-Leitzel, A., Marsh, J., Moynihan, M., & McKelvey, S. (2010). Modeling seasonal influenza outbreak in a closed college campus: Impact of pre-season vaccination, in-season vaccination and holidays/breaks. *PloS One*, *5*(3), e9548.

Pons, V. G., Canter, J., & Dolin, R. (1980). Influenza A/USSR/77 (H1N1) on a university campus. *American Journal of Epidemiology*, *111*(1), 23–30.

Salathe, M., & Jones, J. H. (2010). Dynamics and control of diseases in networks with community structure. *PLOS Computational Biology*, *6*(4), 1–11.

Schwartz, E. J., Morgan, M., & Lapin, S. (2015). Pandemic 2009 H1N1 influenza in two settings in a small community: The workplace and the university campus. *Epidemiology and Infection*, *143*(8), 1606–1609.

Sobal, J., & Loveland, F. C. (1982). Infectious disease in a total institution: A study of the influenza epidemic of 1978 on a college campus. *Public Health Reports*, *97*(1), 66.

Thompson, W. W., Shay, D. K., Weintraub, E., Brammer, L., Cox, N., Anderson, L. J., & Fukuda, K. (2003). Mortality associated with influenza and respiratory syncytial virus in the united states. *JAMA*, *289*(2), 179–186.

Trapman, P., Ball, F., Dhersin, J.-S., Tran, V. C., Wallinga, J., & Britton, T. (2016). Inferring R0 in emerging epidemics—the effect of common population structure is small. *Journal of The Royal Society Interface*, *13*(121), 20160288.

Vaidya, N. K., Morgan, M., Jones, T., Miller, L., Lapin, S., & Schwartz, E. J. (2015). Modelling the epidemic spread of an h1n1 influenza outbreak in a rural university town. *Epidemiology and Infection*, *143*(8), 1610–1620.

Viboud, C., Earn, D. J., Simonsen, L., Dushoff, J., & Plotkin, J. B. (2006). Mortality due to influenza in the united states-an annualized regression approach using multiple-cause mortality data. *American Journal of Epidemiology*, *163*(2), 181–187.

Vynnycky, E., & White, R. (2010). *An introduction to infectious disease modelling*. Oxford: Oxford University Press.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, *466*(7310), 1102–1104.

Zambon, M., Nicholson, K. G., & Wood, J. M. (2003). Influenza. *Lancet (London, England)*, *362*(9397), 1733–1745.