Modelling Covid-19 Transmission at Universities

Emma L. Fairbanks*

Introduction

The first Covid-19 lockdown during March 2020 included closure of universities for most in-person activities, with teaching delivered online for the remainder of the academic year. Over 2.3 million students study at over 160 education providers in the UK. Mass national and international movement at the beginning of academic terms is often associated with illness, commonly referred to in the UK as ‘freshers flu’. When universities reopened their doors in autumn 2020, as Covid-19 cases were increasing, many universities experienced large outbreaks, which led to fears of self-isolation and an increased number of cases *spilling over* into local communities. As well as isolation being detrimental to the mental health of students, the sheer number of isolated students posed logistical problems to the universities, due to the need for food deliveries and other support.

Before the start of the 2020/2021 academic year, there were many unknowns, such as the relationship between Covid-19 prevalence in incoming students and large outbreaks, which factors increased risks in student halls, the impact of university outbreaks on surrounding communities, and the impact of isolation, testing and adherence on the progression of Covid-19 in university settings.

Despite these uncertainties, decisions had to be taken on the operation of universities during the pandemic. Mathematical modelling was a valuable tool in the decision-making process. This article summarises the research by the Higher Education working group at the Isaac Newton Institute for Mathematical Sciences, University of Cambridge, published in [1], CC BY 4.0[1].

First, we will discuss a retrospective analysis that sought to understand what occurred during the autumn term of the 2020/2021 academic year, attempting to answer the questions:

- Can the estimated numbers of incoming infectious students explain the observed outbreaks?
- What factors influence infection risk in student halls?
- How do university outbreaks impact local communities?

Then, we discuss models used to analyse questions associated with potential future scenarios if students were to return to university:

- Does adhering to test, trace and isolation rules reduce the probability of infection?
- Does mass testing of students reduce the probability of an outbreak?
- If a more transmissible variant were to circulate, would testing need to be increased?

Additional analysis may be found in the source paper, as well as a much more rigorous criticism of the assumptions and limitations of the analysis, which, of course, need to be considered when providing evidence to policymakers! The whole body of work was presented as evidence to inform government decision-making when the government road map to lift the lockdown was announced on 22 February 2021.

Do numbers of incoming infectious students explain the observed outbreaks?

The universities had many rules to try to prevent transmission; however, many still experienced large outbreaks. With such a large number of students moving around the country and the number of UK cases increasing, were these outbreaks to be expected?

If we define p as the probability that an incoming infection fails to cause an outbreak, then p^n is the probability that n incoming infections will fail to cause an outbreak. If n is the initial number of infectious students, then the probability of an outbreak is $\mathcal{P} = 1 - p^n$. If $p = 1$, there is no probability of an incoming infection causing an outbreak. Therefore, the probability of a university outbreak is $\mathcal{P} = 1 - 1^n = 0$ for all possible values of n. In contrast, if $p = 0$, all incoming infections would be expected to cause a university outbreak, and therefore, the probability of a university outbreak is $\mathcal{P} = 1 - 0^n = 1$ if n is larger than 0 and $\mathcal{P} = 1 - 0^0 = 0$ if $n = 0$ (since $0^0 = 1$).

Enright [2] estimated the number of incoming infected students for each university using data on the proportion of communities testing positive (prevalence) and data on students’ home and term-time postcodes. To learn about the size of the observed outbreaks in university settings, we collected outbreak data from the University and College Union (UCU) dashboard [3] for 72 universities.

Box 1: Maximum likelihood estimation

Maximum likelihood estimation calculates the likelihood function (the likelihood of the data given the model) for a range of values of p. Therefore, it finds the value of p that best describes the data. For each university i, we represent the imported number of cases by n_i. We set $x_i = 1$ if a university experienced an outbreak and $x_i = 0$ if not. The likelihood of the data for all N universities given p is then

$$L(p) = \prod_{i=1}^{N} \underbrace{p^{n_i(1-x_i)}}_{\text{no outbreak}} \underbrace{(1-p^{n_i})^{x_i}}_{\text{outbreak}}.$$

Here, if $x_i = 0$, then $(1 - p^{n_i})^{x_i} = 1$, and therefore, $p^{n_i(1-x_i)}(1 - p^{n_i})^{x_i} = p^{n_i}$, i.e. the probability that all n_i cases do not lead to an outbreak. If $x_i = 1$, then $p^{n_i(1-x_i)} = 1$, and as a consequence, $p^{n_i(1-x_i)}(1 - p^{n_i})^{x_i} = 1 - p^{n_i}$.

There is always uncertainty around statistical estimates because an estimate is based on the sample of the population analysed. If a different sample of the population were analysed, the range of values we expect the estimate to fall into for a given percentage of samples is called the confidence interval.

---

*Swiss Tropical and Public Health Institute and University of Basel, Switzerland

Table 1: Expected impact of increasing hall capacity and proportion of students sharing a bathroom on the hall SAR (95% confidence intervals) from the final multivariate logistic regression (from [1]).

	Percentage of students without a private bathroom		
Hall size	**0%**	**50%**	**100%**
100	0.06 (0.04–0.08)	0.07 (0.06–0.09)	0.09 (0.07–0.11)
200	0.08 (0.07–0.10)	0.10 (0.09–0.12)	0.13 (0.11–0.14)
400	0.16 (0.13–0.19)	0.19 (0.17–0.22)	0.23 (0.20–0.27)

Universities were binned, with bin widths of 10, according to the estimated number of incoming infectious students. For each bin, we calculated the fraction of universities that experienced a large outbreak, which we subjectively defined as 400 or more reported cases by 19 November 2020. The model estimates followed the trend of the observed fraction of universities experiencing an outbreak (Figure 1).

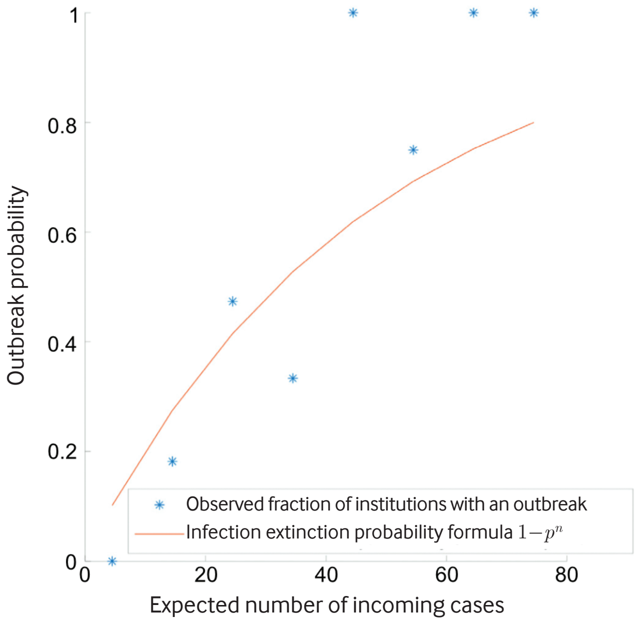

Figure 1: Outbreak probability as a function of expected incoming infections, where p is fitted by maximum likelihood estimation (from [1]). See Box 1.

Generally, estimates for p were relatively large and close to 1, indicating that most chains of infection died out without causing a large outbreak. This could suggest that the universities' strategies for preventing outbreaks worked quite well. However, if the sheer number of incoming cases is large (due to the large number of cases across the UK), more large university outbreaks should be expected. The value of p not only depends on the transmissibility of the variant and the rules in place at the time, but also on how well students follow them.

What influences infection risk in student halls?

Many students had to isolate due to Covid-19 infection in halls of residence. Universities attempted to prevent this by splitting halls into households, which were intended to function similarly to households in the community. There were restrictions on socialising beyond household members and requirements for the entire household to isolate if a member displayed symptoms or tested positive. Factors influencing the risk of infection amongst students in halls at a single university were examined using multivariate logistic regression (see Box 2). Our analysis was performed on reported cases, with actual infection levels likely higher!

We considered an event to be a student in a hall testing positive for Covid-19 given at least one other student in the hall had tested positive, referred to as the secondary attack rate (SAR). The predictors we included in the model were:

- Median household size.
- Proportion of students in medical courses, as a proxy for higher risk of infection due to attending placements.
- Hall size.
- Proportion of students sharing a bathroom.

Only hall size and proportion of students sharing a bathroom were significantly associated with SAR. We provide the predicted SAR for different hall capacities and proportions sharing bathrooms in Table 1. The data indicate that students in halls with all shared bathrooms (final column) are approximately 50% more likely to become infected than students in halls with all en-suite rooms (first column). Our results suggest that dividing students into households was unlikely to reduce infection risk. In contrast, the model suggested that partially filling halls could significantly reduce transmission risk, especially if this also reduced shared spaces.

Box 2: Multivariate logistic regression

Multivariate logistic regression is a statistical model of the probability of an event (p), given as

$$p(x_1, x_2, \ldots, x_n) = \frac{1}{1+\mathrm{e}^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n)}},$$

where x_1, x_2, …, x_n are predictors and β_1, β_2, …, β_n are the corresponding parameter values of the impact the predictor has on the probability of the event. For each predictor, a p-value, indicating the statistical significance of how well the model fits the data when it does and does not include the predictor, is also calculated. If a p-value is less than 0.05, we say it is significant.

How do university outbreaks impact locals?

Asking this question another way, following a student outbreak, could *spillover* signals (more community cases than expected) be detected? A spillover signal likely depends on:

- The size of the student outbreak.
- The number of cases in the community at the start of the outbreak.
- The proportion of local students (who attend a university in the same region as their home address).

To investigate, we decided to examine the temporal patterns of confirmed cases in each university and local community.

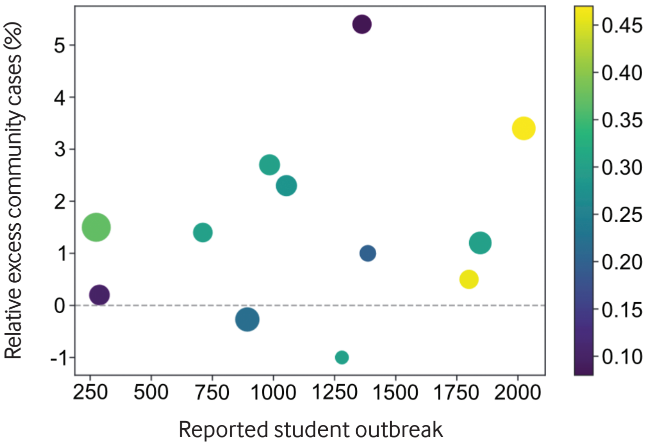

Figure 2: Relationship between student cases and local community cases, with the proportion (15% to 77%) of local students (circle size); community incidence per 1000 at the time of peak student cases (colour); potential spillover signal above the dashed line, and no signal below it (from [1]).

We collected data for students from the UCU Covid-19 dashboard [3]. Collecting relevant data for community cases was more tricky! Though data were available by age from Public Health England, which has since been replaced by the UK Health Security Agency and the Office for Health Improvement and Disparities, these data were likely also to include the student cases. Therefore, as a proxy for community cases, we used the number of cases for people who were not of the typical student age (18–24).

We detected diverse levels of potential spillover, shown in Figure 2. Some of the largest university outbreaks were in areas with the largest number of community cases (yellow dots on the right). Larger university outbreaks tend to correlate with a larger spillover signal (dots in the top right corner). However, this is not always the case (dots in the lower right corner). Some universities had large outbreaks but only a small spillover signal could be detected. Sometimes, there was no spillover signal at all.

This analysis shows how complex the issue is; there is not always a clear relationship between the size of a university outbreak and the number of community cases due to spillover! In practice, the spillover could be from the local community to students, if the community has a high enough prevalence.

Here, the definition of a spillover assumes that there is a causal relationship between student and community cases, with no other factors. Many other factors were likely to influence community prevalence; for example, children returned to schools at a similar time as students returned to university.

Does adhering to test, trace and isolation rules reduce the probability of infection?

To minimise the risk of large outbreaks during the second academic term, guidance stipulated that all students should be offered a test when they returned to university. This helped to identify and isolate asymptomatically infected individuals [4]. The testing regime involved two lateral flow tests, three days apart, with isolation between tests.

Although there are many different types of model for simulating disease transmission, to investigate the extent that adherence to test, trace and isolate measures can minimise the impacts of outbreaks, we needed a model framework that can capture the individual nature of the intervention. We, therefore, used a network model.

Network models consider contact patterns as graphs, where the nodes represent hosts and the edges represent contacts. Each contact with an infectious host is associated with a probability of infection. Figure 3 is a schematic of an example outbreak.

If transmission takes place, each newly infected individual enters a latent period of infection (when they are not yet infectious). Then, when they are infectious, they either remain asymptomatic (an asymptomatic case) or transition through the presymptomatic and symptomatic stages of infection. Asymptomatic testing can detect both presymptomatic and asymptomatic cases. When individuals are identified by asymptomatic testing or display symptoms and enter isolation, they do not continue to infect others.

At the time we were conducting our analyses, the sensitivity of lateral flow tests was an area of considerable uncertainty.

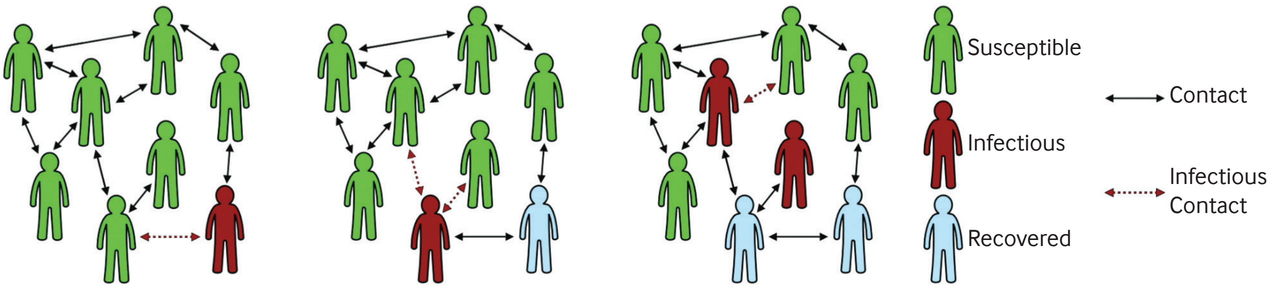

Figure 3: Example of a network model for an outbreak.

During the initial stages of infection, when the virus is proliferating, it was assumed sensitivity was the same for symptomatic and asymptomatic individuals. After this, during the clearance stage, it was assumed the probability of asymptomatic individuals testing positive decayed faster (Figure 4).

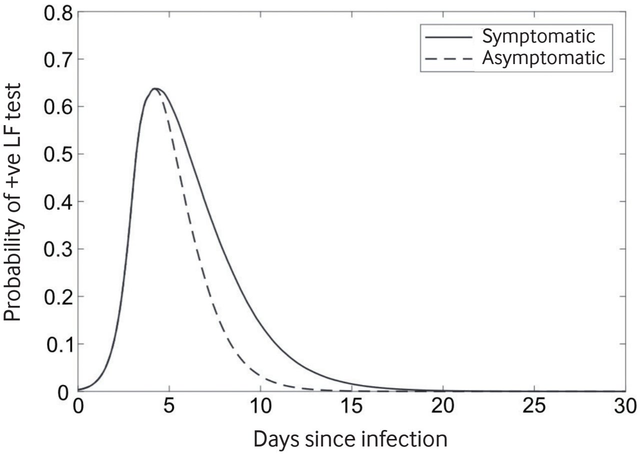

Figure 4: Probability of symptomatic and asymptomatic individuals having a positive lateral flow test result (from [1]).

We considered a network of 25 000 students, with 7155 on campus and the remainder off campus. Students were assumed to have contact with all household members each day. Using data from a student social contact survey conducted in 2010 [5], students were randomly assigned non-household contacts (such as with their study cohort, organised societies and sports clubs, and other social contacts). These data were used to build 50 unique possible networks. Each network was then simulated 20 times, resulting in a total of 1000 simulations. In each simulation, a background (community) prevalence was sampled from a uniform distribution, with minimum 0.005 and maximum 0.02, and this was used to compute a daily probability of infection for off-campus students. Each simulation was run for the duration of a term (11 weeks). For further details about this model and its limitations, see [6] for a full description.

The model was used to simulate a baseline scenario (with no interventions) and intervention scenarios (with no society contacts and on-campus students adhering to the household bubble rules within halls). Intervention scenarios also considered different probabilities of students adhering to testing on return, isolation (10 days for contacts of confirmed cases) and contact testing.

Even when no students were adherent (an adherence probability of 0), contacts were reduced due to the lack of society activities and household bubbles within halls, and therefore, the relative attack rate was reduced compared to the baseline scenario (Figure 5).

Nevertheless, the analysis shows adherence is crucial in reducing the overall number of cases! Initially, as more students adhere to the rules, the average time in isolation increases (as expected). However, the proportion of time in isolation per student peaks when roughly 70–80% of students adhere, with time in isolation per student being lower if all students adhere. We observe this effect because the increased isolation due to adherence is outweighed by the reduction in the number of cases.

Does mass testing of students reduce risk?

Between November 2020 and January 2021, the Alpha variant spread across the UK. It was presumed to be more transmissible than previously widespread variants. This motivated interest in whether mass testing could control this more transmissible variant. Would testing regimes that would have worked in the previous term still be sufficient if students were to return to university?

Asymptomatic testing systems were also used throughout the term. A similar network model was used to simulate contacts of 15 000 students with household and non-household members. Again, infected individuals transitioned through a latent period before becoming asymptomatic or presymptomatic and then symptomatic. In this model, it is was assumed that half of infected individuals remained asymptotic and that all infectious individuals were equally infectious.

As well as a baseline scenario with no testing, we explored four asymptomatic testing scenarios in which individuals were tested at random with a probability $1/3$, $1/7$, $1/10$ or $1/14$ per day per person (to simulate testing every 3, 7, 10 or 14 days, respectively). In all scenarios, it was assumed that when an

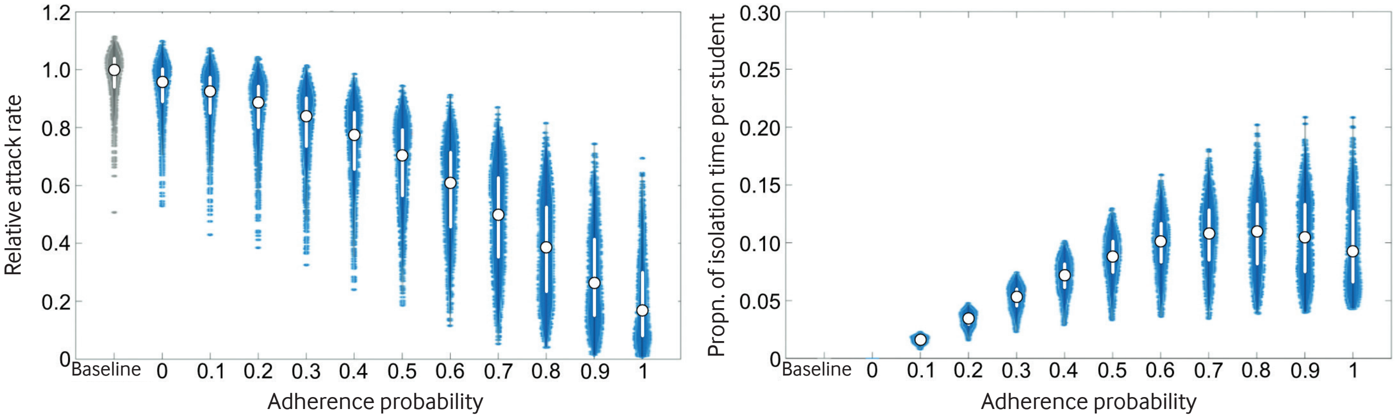

Figure 5: Output of simulations for various probabilities of adherence given all students participate in testing when returning to university. Distributions correspond to: relative attack rate compared to the baseline scenario (left) and time spent in isolation per student (right). The white markers denote medians and solid black lines span the 25th to 75th percentiles (from [1]).

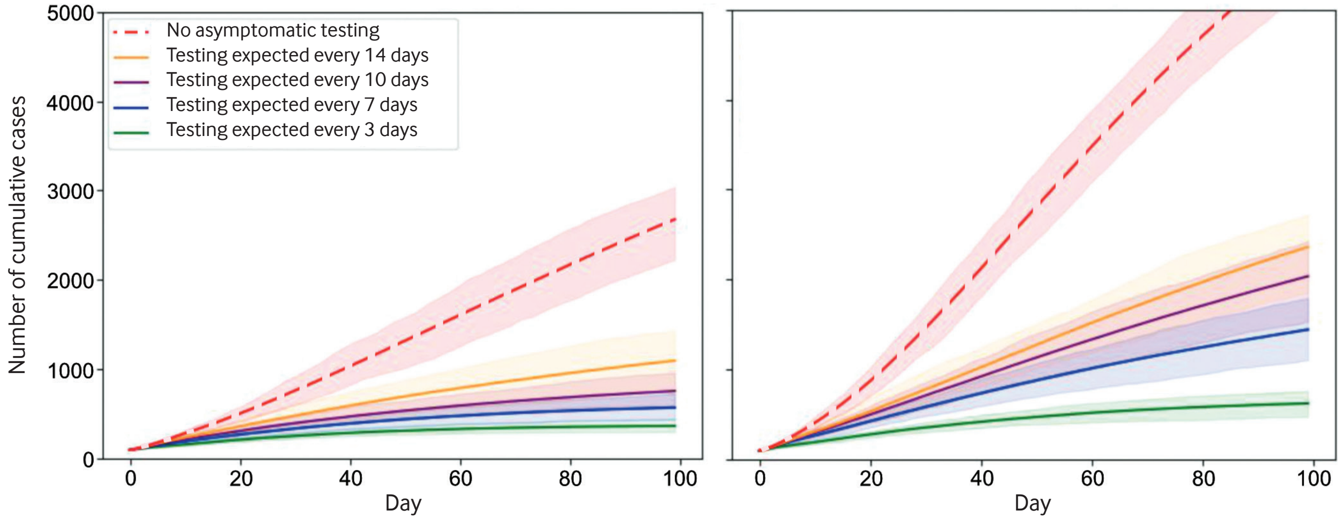

Figure 6: Time series of cumulative cases for each testing scenario for the lower-transmissibility variant (left) and the higher-transmissibility variant (right). Shaded regions show the 95% prediction intervals, and solid lines show mean values (from [1]).

individual became symptomatic they isolated immediately. If an individual tested positive, then their entire household isolated for 14 days.

Simplifying assumptions included perfect testing (all infected individuals receive a positive result) and an adherence probability of 1. Here, an adherence probability of 1 corresponds to all individuals with a positive test or symptoms isolating, as well as their household members. It was also assumed that 50% of non-household contacts were traced and isolated.

The model was analysed for a lower-transmissibility variant (intended to simulate the variant circulating in autumn 2020) and a higher-transmissibility variant, 1.5 times more transmissible. This was implemented by multiplying the probability of transmission due to a contact with an infectious individual by 1.5.

The model was run for 100 days starting with 100 students randomly assigned as infectious. The model was simulated for 100 replicates, each run on a newly generated network. Figure 6 shows that more frequent testing leads to less cases overall. The figure suggests that if each student participated in testing once every two weeks then the number of cases would be approximately halved. Many universities offered testing twice a week (most similar to once every three days). In this model, if all students followed this advice, the number of cases would be significantly reduced compared to students choosing not to test.

Conclusions

These mathematical investigations produced the following findings:

1. High community prevalence leads to many infected incoming students, increasing the probability of large outbreaks.
2. Living in large halls with shared facilities poses a high risk of infection.
3. There is no clear relationship between student outbreaks and an increase in local community cases.
4. Adhering to the rules decreases the attack rate.
5. To avoid large outbreaks, universities have to test students very frequently, especially if there are high-transmissibility variants.

Epidemiological analyses, such as this, are still ongoing to help protect and improve the well-being of students and the wider community!

Notes

1. This research [1] was published by the Royal Society under the terms of the Creative Commons Attribution License (`creativecommons.org/licenses/by/4.0/`).

REFERENCES

1 Enright, J. et al. (2021) SARS-CoV-2 infection in UK university students: lessons from September–December 2020 and modelling insights for future student return, *R. Soc. Open Sci.*, vol. 8, no. 8, art. 210310.

2 Enright, J. (2020) Basic arriving student calculations, `tinyurl.com/Enright-data2020`.

3 UCU (2020) Monitoring cases of Covid-19 in UK higher and further education, `tinyurl.com/webarchive-dashboard`.

4 Department for Education (2020) Press release: All students offered testing on return to university, `www.gov.uk/government/news/all-students-offered-testing-on-return-to-university`.

5 Danon, L. et al. (2012) Social encounter networks: collective properties and disease transmission, *J. R. Soc. Interface*, vol. 9, no. 76, pp. 2826–2833.

6 Hill, E.M. et al. (2021) Modelling SARS-CoV-2 transmission in a UK university setting, *Epidemics*, vol. 36, art. 100476.