

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/136640/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Harper, Paul, Moore, Joshua and Woolley, Thomas 2021. Covid-19 transmission modelling of students returning home from university. *Health Systems* 10 (1), pp. 31-40. 10.1080/20476965.2020.1857214 file

Publishers page: <https://doi.org/10.1080/20476965.2020.1857214>  
<<https://doi.org/10.1080/20476965.2020.1857214>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Covid-19 Transmission Modelling of Students Returning Home from University

Paul R. Harper, Joshua W. Moore, and Thomas E. Woolley

Cardiff School of Mathematics Cardiff University Senghennydd Road, Cardiff, CF24 4AG, United Kingdom

November 26, 2020

## Abstract

We provide an open source model to estimate the number of secondary Covid-19 infections caused by potentially infectious students returning from university to private homes with other occupants. Using a Monte-Carlo method and data derived from UK sources, we predict that an infectious student would, on average, infect 0.94 other household members. Or, as a rule of thumb, each infected student would generate (just less than) one secondary within-household infection. The total number of secondary cases for all returning students is dependent on the virus prevalence within each student population at the time of their departure from campus back home. Although the proposed estimation method is general and robust, the results are sensitive to the input data. We provide Matlab code and a helpful online app ([http://bit.ly/Secondary\\_infections\\_app](http://bit.ly/Secondary_infections_app)) that can be used to estimate numbers of secondary infections based on local parameter values. This can be used worldwide to support policy making.

## 1 Introduction

SARS-CoV-2, otherwise known as Covid-19, or corona virus, has caused a global pandemic lasting throughout 2020 (Wang et al. 2020, Chahrour et al. 2020, Sahu 2020). Governments tried to stem the infection spread by imposing national lockdowns, resulting in many services, such as education, rapidly transitioning to online spaces (Rapanta et al. 2020).

Concerns had been raised (Yamey & Walensky 2020, Perez-Reche & Strachan 2020) about potential high rates of transmissions between university students as they were encouraged to return to campus, especially those living together in large capacity, close confinement, housing, such as university halls of residence. Indeed, many universities initially reported localised surges in infections (Teran et al. 2020) confirming indications that infections are more likely to occur due to contact with peers rather than across age bands (Laxminarayan et al. 2020).

It could be suggested that since young people (the typical age of a student) generally recover well from Covid-19 (WHO 2020) then allowing them to pass the disease rapidly between themselves is actually a good way of producing herd immunity, at least for their age group (Britton et al. 2020). However, questions regarding how long the immunity will last (Ripperger et al. 2020, Tan et al. 2020, Callow et al. 1990, Seow et al. 2020) coupled with suffering through the actual disease and the stress of missing coursework (Husky et al. 2020) leads this to be an unsympathetic plan of action.

Critically, even if herd immunity was possible in the university student body the students are going to return to their private domiciles at the end of term and pose a risk to everyone in their household. Although there has been a focus on modelling the future trajectories of an infection outbreak within a university (Christensen et al. 2020), our work explores an area that has seemingly not been previously considered: namely, what is the impact if students return home during lockdown, or at the end of term? Even if students are able to recover well, or even be asymptomatic, they can still be carriers (Laxminarayan et al. 2020). Thus, they can take the infection into a household that may contain other adults of different age groups (*e.g.* parents) that may not be expected to recover as well, particularly if they have an underlying long term illness. Further, this is a timely question to ask due to the winter holiday break being a major time for family gatherings resulting in a higher risk of extended contacts with cross-sectional age groups.

Due to the rapid, flexible and predictive nature of mathematical modelling, mathematicians, statisticians and operational researchers have been at the very heart of understanding the evolution of the pandemic (*e.g.* Ferguson et al. (2020)) and ways to reduce the impact (*e.g.* Currie et al. (2020)), leveraging a variety of different techniques including deterministic and stochastic differential equations (Ferguson et al. 2020), machine learning (Latif et al. 2020), simulation (Mahmood et al. 2020, Currie et al. 2020) and queueing theory (Zhigljavsky et al. 2020).

Our contribution is to apply Monte-Carlo theory to UK-based data and thereby estimate the number of secondary infections that will arise from university students returning home. Namely, we combine the prevalence of infection at university (in this case Cardiff University), with distributions of household size and probability of secondary infections, to predict the mean expected number of secondary infections, as well as 95% confidence intervals.

Although our results are based on UK data we provide the reader with Matlab code (Appendix A) and an online app ([http://bit.ly/Secondary\\_infections\\_app](http://bit.ly/Secondary_infections_app)) that not only reproduces our results, but can also be adapted by the reader to include data which is more accurate and/or specific to their location and needs. The full Matlab and app codes are kept in an online repository ([http://bit.ly/secondary\\_infections\\_repo](http://bit.ly/secondary_infections_repo)) and will be updated as required. The user's local data and subsequent results can be used to develop and evidence a university's strategy for dealing with an infected student body and their desire to return home for the holidays.

Our results have been presented to TAG (Task Advisory Group for Welsh Government Covid-19 response) and the Wales Higher Education Covid-19 Task and Finish Group, and was used to inform policy in relation to the two-week firebreak (lockdown) in Wales during the period 23 October - 8 November, 2020, when students were asked to remain at their university lodgings, rather than return home. The data has also been communicated across the governments of England, Scotland and Northern Ireland, so that it can inform the wider development of policy development of planning for the winter holidays and beyond.

Section 2 presents the fundamental equation that allows us to estimate the number of secondary infections that arise from infected students returning home. To extend beyond a single point estimate of the equation we introduce the Monte-Carlo algorithm in Section 2.1, which allows us to estimate the variability of our prediction. Critically, although the Monte-Carlo algorithm is a robust method for estimating such values our results do depend on statistics from various sources, thus, sections 2.2-2.4 are focused on presenting and evaluating this data. Our numerical codes and online app are discussed in Section 3, where we provide a brief manual for the online app, demonstrating its complete generality with regard to alternative input data. The resulting numerical predictions are presented in Section 4 and discussed in Section 5. Finally, we note in Section 5 the work presented is not only useful for predicting secondary infections that occur due to students journeying back home it can also be adapted to estimate the secondary infections that will initially arise when students return to university.

## 2 Methods and data sources

We seek to define an equation that will provide a prediction of the number of secondary household infections arising from students returning home from university. Critically, our proposed equation is simple, thus, the reader will have a clear understanding of what information is required and how it is combined. Further, the equation combines parameters that are estimatable from known data. Thus, although we provide values, which are specifically useful for Welsh Government policy making, we provide an app and adaptable code that can be used with data that is more current and more localised to a user’s requirements.

To generate our proposed equation we first define a number of variables. Specifically, we define:

- $I$  to be the prevalence of the virus in the student body. Specifically, this is the percentage probability that an individual student has Covid-19 and is infectious. In this study this is as a control parameter, which is informed by data from (ONS 2020, Cardiff University 2020). See Section 2.2.
- $S$  to be the probability of a secondary transmission to another member of the household. This is a continuous random variable, assumed to be normally distributed. The mean and standard deviation is informed by data from (Lopez Bernal et al. 2020). See Section 2.3.
- $H$  to be the number of occupants in a household, other than the student themselves. This is a discrete random variable whose distribution is informed by (Welsh Government 2020). See Section 2.4.
- $N$  to be the total number of students returning back home from campus. Here, we take  $N = 1000$ , so the results can be stated “per thousand students”.

The total number of additional (secondary) household infections,  $T$ , is thus calculated by:

$$T = \underbrace{\frac{I}{100} \times N}_{\substack{\text{Expected number} \\ \text{of infected} \\ \text{students}}} \times \underbrace{S \times H}_{\substack{\text{Probability of} \\ \text{secondary} \\ \text{infection given} \\ \text{infected} \\ \text{household} \\ \text{member}}} . \quad (1)$$

Although using the average values of  $S$  and  $H$  would provide a “rule of thumb” measure for  $T$ , we have decided to extend this investigation and use a Monte-Carlo approach to sample from the specified distributions of  $S$  and  $H$  (Landau & Binder 2014). Not only does this method allow us to capture the stochastic nature of virus transmission, but equally, it will provide us with error bounds on the number of predicted secondary infections.

### 2.1 Monte-Carlo algorithm

Monte-Carlo algorithms are used in a wide variety of fields such as: quantum mechanics, statistical physics, telecommunications, transportation, biology and medicine (Landau & Binder 2014, Rubinstein & Kroese 2016, Fishman 2013, Manly 2006). The strength of the Monte-Carlo approach lies in its ability to forecast not just an estimate for the mean of the value that is being calculated, but provide error bounds around this estimate, providing the user with confidence intervals as to their uncertainty.

Monte-Carlo algorithms are particularly applicable to our question as we can take advantage of the distribution information that is provided, rather than just the mean values. For example, if we ignore the single person households then the data in Section 2.4 suggests that the mean household size is approximately 3.4 people. So, on average, there are approximately 2.4 other people in the household that an infected student could infect. Such a simple description of household size is problematic though, since it takes the discrete size of a household and makes it a continuum, which does not make sense for an individual household.

Further, the single point estimate does not provide a sense of the spread of data. Namely, even though two data sets may have the same mean value a data set with a high variance may be less reliable than a data set with a small variance. Thus, the Monte-Carlo approach combines the uncertainty of the input data and provides a measure of uncertainty in the output.

The central idea of the Monte-Carlo algorithm is to sample many times (specifically  $N$  times) from the household size and secondary infection probability distributions to generate approximations for  $H_k$  and  $S_k$   $k = 1, \dots, N$ , respectively. This removes the problem of assuming continuous mean values for discrete distributions and allows us to predict the variability through considering the variance of the  $N$  simulated values.

We present a general method for sampling from these distributions, which can be coded up in a users' language of choice. Alternatively, we provide a full working Matlab code in Appendix A, or for those who want a quick result without resorting to coding we offer an online app, which can be accessed at [http://bit.ly/Secondary\\_infections\\_app](http://bit.ly/Secondary_infections_app).

The household size probabilities form a discrete distribution. Specifically, we define  $p_i$  to be the probability of there being  $1 \leq i \leq n$  people in the household other than the student. For clarity, we explicitly state that  $p_i$  would be the proportion of households containing  $i + 1$  people, *i.e.* the infected student and the  $i$  'other' members of the household. We note that in our current case  $n = 5$ .

To sample from this distribution we generate a uniformly distributed random number,  $r$  and the susceptible household size, is defined to be  $j$  such that

$$\sum_{i=1}^{j-1} p_i < r \leq \sum_{i=1}^j p_i, \quad (2)$$

where we note that  $p_0 = 0$  because we are only considering the cases where students are returning to households containing potential susceptible people. This will lead to a small over estimate as we are ignoring any student who lives on their own outside of term time, however, we assume such cases are in a small minority. Finally, we denote  $H_k = j$  to be the  $k$ th family size.

From the data on secondary infections (Lopez Bernal et al. 2020) we are supplied with the mean probability of secondary infection,  $\mu_i$ , and 95% confidence intervals for households of size  $i = 2, 3, \dots, 6$ . We assume that the probability is Normally distributed and convert the 95% confidence interval back into a standard deviation,  $\sigma_i$ . We then sample  $N$  times from these Normal distributions,  $\mathcal{N}(\mu_i, \sigma_i^2)$ , for each family size,  $i$ , capping the probability at zero and one. Finally, we denote  $S_k(i)$  to be this sampled value, *i.e.* the  $k$ th probability of secondary infection in a household of size  $i$ .

Finally, we simulate  $N$  Bernoulli trials with probability of "success" given by  $I/100$  and defining  $I_k = 1$  if the simulation is "successful" and  $I_k = 0$  if not. Here,  $I_k$  is acting as an indicator variable that the  $k$ th household is home to an infected student.

The total number of secondary infected cases is then given by

$$T = \sum_{k=1}^N I_k S_k(H_K) H_k. \quad (3)$$

The above procedure can be repeated as many times as required to resolve the statistics to any required accuracy. Further, it can be observed that in the case where we collapse the distributions onto their mean values equation (3) simplifies to equation (1).

## 2.2 Probability that an individual student has Covid-19 and is infectious when returning home, $I$

Obtaining an accurate value of  $I$  is problematic due to it evolving over time and varying across the UK. Specifically,  $I$  is influenced by many factors, such as local and national lockdown measures and adherence to social distancing. To include such variance we present the total number of new infections,  $T$ , for a range of different values of  $I$  that reflect a range of localised prevalences possible at the time of departure (when students return home).

Recent ONS data (ONS 2020) suggests that for the age band 12-24,  $I = 1.5\%$  *i.e.* 1.5% of those in that age category are infected. However, case data from Cardiff University’s asymptomatic testing service (Cardiff University 2020) indicated that the underlying rate may have been as high as  $I = 15\%$  (w/c 10th October 2020) but at the time of writing the rate has dropped to well under  $I = 5\%$ . To encompass this range we use  $I \in \{0.5, 1.5, 5, 10, 15\}$ .

## 2.3 Secondary transmission probabilities, $S$

Several publications, for example (Wu et al. 2020, Rosenberg et al. 2020, Lewis et al. 2020) have reported estimates for  $S$ . We have used those in a more recent UK-wide study (Lopez Bernal et al. 2020) given the UK context and that, reassuringly, the estimate is broadly in-line with other studies. Overall, the authors report the probability of a secondary infection  $S = 0.37$  *i.e.* 37% (95% CI 31-43%). In this Monte-Carlo study we use the more detailed information from (Lopez Bernal et al. 2020), where the authors are able to differentiate values of  $S$  by household size. The data set can be found in Table 1.

Household size	Probability of secondary infection with 95% confidence interval
2	0.49 (0.37 - 0.60)
3	0.41 (0.29 - 0.52)
4	0.32 (0.22 - 0.42)
$\geq 5$	0.25 (0.14 - 0.36)

Table 1: Probability of secondary infections occurring in households of different sizes. Adapted from Lopez Bernal et al. (2020).

We note that in households of size 2,  $S = 0.49$  (95% CI 0.37 – 0.61) *i.e.* a Covid-19 positive student has, on average, a 49% chance of making the other occupant infected. However, for houses with 5 or more occupants, this reduces to  $S = 0.25$  (25%) for each occupant. It is unexpected that the probability should decrease with an increased household size. To account for this, we suggest that this may be an actual result that shows that larger households are more likely to be more cautious in their interactions. Alternatively, this may simply be a problem with the low number of recordings in the data. Namely, Covid-19 is, thankfully, rare and household sizes  $\geq 5$  are also rare (see Section 2.4), thus the  $S = 0.25$  value is derived from only  $\sim 30$  reported cases of secondary infections occurring in such households.

## 2.4 Household size, $H$

Data on the number and proportion of households in Wales containing at least one student (2019) has been provided by Economic and Labour Market Statistics, Welsh Government, and is summarised below. The average household size can be calculated to be 3.2 people, however, we use the full distribution provided by the data in the Monte-Carlo simulation, as described in Section 2.1. Specifically, each household size is linked to different values of the probability of secondary transmissions,  $S$ .

Number of people (including students) in a household	Number of households with at least one student	Proportion of all households containing at least one student (%)
1	8,300	8.3
2	22,000	22.1
3	27,600	27.6
4	27,700	27.7
5	10,100	10.1
6+	4,300	4.3
Total	<b>99,900</b>	<b>100</b>

Table 2: Number and proportion of households in Wales containing at least one Higher Education student by household size. Source: Welsh Government analysis of the Annual Population Survey, 2019 (Welsh Government 2020).

## 3 Codes and app

Alongside the Matlab code in Appendix A that can be used to reproduce our results, as well as be adapted to a user’s specific data set, we have also developed an online app, which can be found at [http://bit.ly/Secondary\\_infections\\_app](http://bit.ly/Secondary_infections_app), for ease of use. Full Matlab and app codes will be kept in an online repository, [http://bit.ly/secondary\\_infections\\_repo](http://bit.ly/secondary_infections_repo).

Figure 1 illustrates the user interface. The top contains a table of values which can be edited by the user. The values in the table must all be positive and less than one. Further, the sum of the probabilities along the top ‘household’ row, must sum to less than one. Next the sliders on the left are set to the chosen number of students to be simulated,  $N$ , and the chosen prevalence rate,  $I$ .

Once these values are set, the ‘Run’ button below the table on the right should be clicked and after a few seconds of simulation the results table will be filled with predicted results. Namely, the top row corresponds to the user’s chosen prevalence rate and presents the mean number of secondary infections, as well as the 95% confidence intervals for the prevalence rates. The rows below this present the prevalence rates for  $I \in \{0.5, 1.5, 5.0, 10.0, 15.0\}$ , which are used to compare with the results presented in this paper.

## 4 Results

Every time we sample from the probability distributions defined by the data, we generate a single sample of  $T$ . Some values of  $T$  will be more common than others because some values of  $S$  and  $H$  are more common than others (see sections 2.3 and 2.4). Creating a histogram of these samples elucidates which values of  $T$  are most common and how they are spread over their possible values.

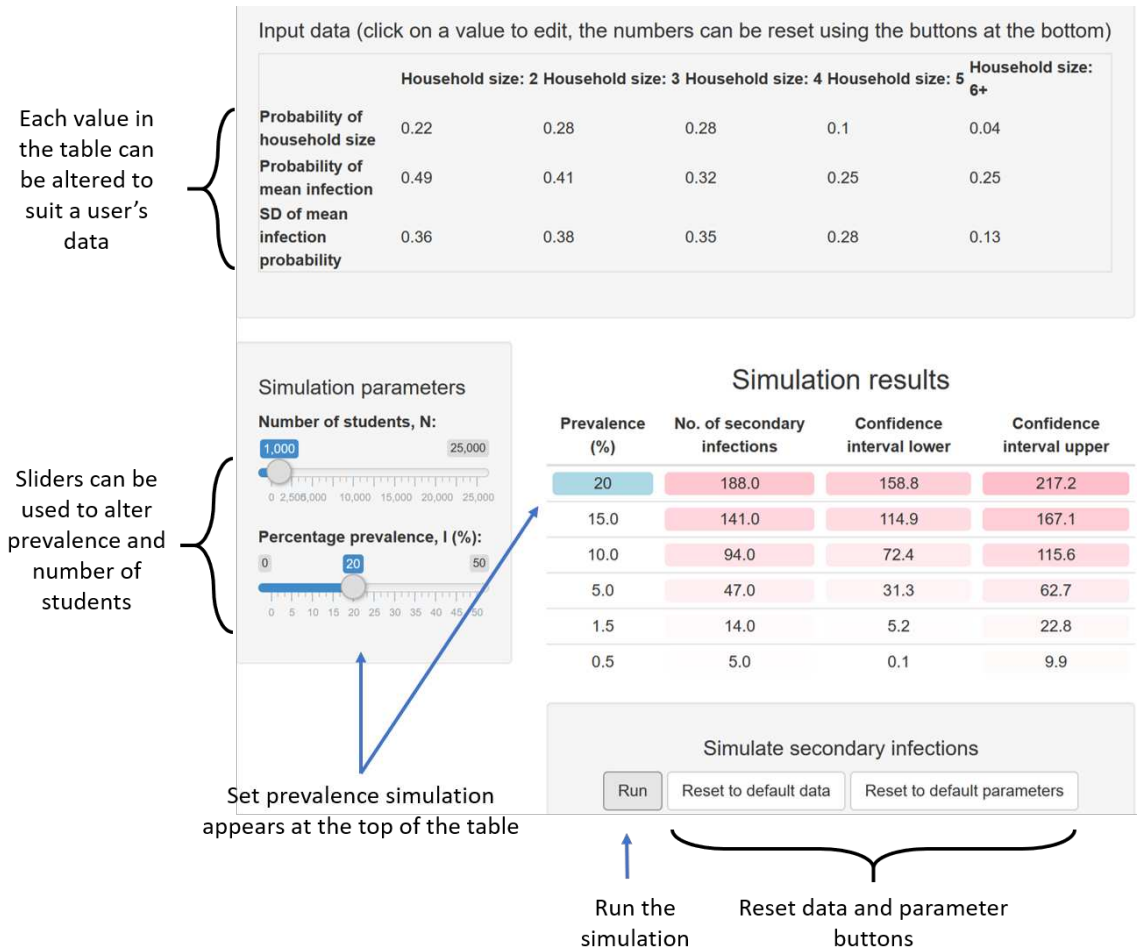


Figure 1: Online app interface. The app can be found at [http://bit.ly/Secondary\\_infections\\_app](http://bit.ly/Secondary_infections_app).

From these histograms we are able to estimate the mean value and accompanying 95% confidence intervals of  $T$ .

Critically, quoting only the mean and 95% confidence interval statistics suggests that we are assuming that our underlying distribution is Gaussian, or Normal. To help the reader understand how good this assumption is we fit a Gaussian distribution to the sampled data. A good fit between curve and histogram means that the data is closer to being Normally distributed and, thus, our statistics will provide a better estimate for the data spread.

Figure 2 presents histograms of  $10^4$  Monte-Carlo stochastic trials calculating  $T$  from equation (3), for varying prevalence rates. Specifically, figures 2(a)-2(e) represent prevalence levels of  $I = 0.5\%$ ,  $1.5\%$ ,  $5\%$ ,  $10\%$  and  $15\%$ , respectively. Due to the central limit theorem, the assumption that the final distribution is Gaussian is more applicable to the simulations with higher prevalence rates, *i.e.* when more people are infected. However, even in the low prevalence case of  $I = 0.5\%$  the Gaussian fit is easily within the error tolerances of the data we have.



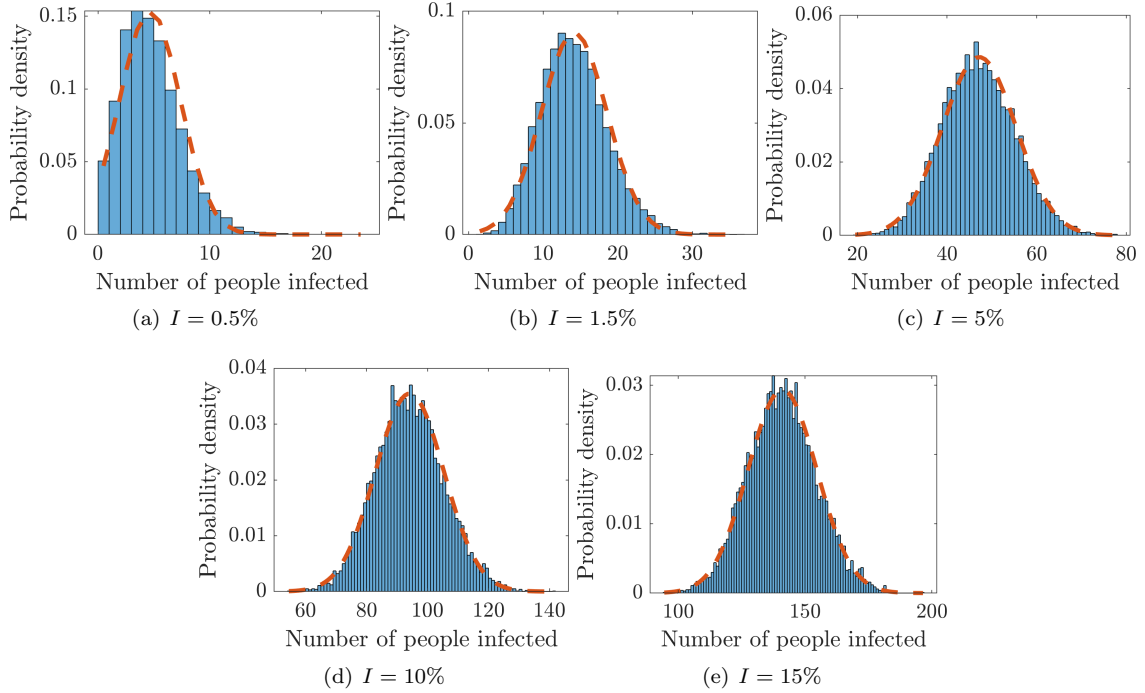


Figure 2: Histogram and data fit representing the variability over  $10^4$  Monte-Carlo simulations. The blue histogram is the simulated data, the red dashed line is the fit of a normal distribution. Note that we are plotting the probability density, but since each column is one person wide the  $y$  values can be interpreted as a probability. The prevalence,  $I$ , is noted below each figure.

Estimates for the mean value of  $T$ , over the different  $I$  values, as well as 95% confidence intervals can be derived from these distributions and are collected in Table 3. Results are displayed for a value of ‘per 1000 students’ *i.e.*  $N = 1000$ .

Considering the average number of secondary cases, we notice that a rough ‘rule of thumb’ for the expected number of secondary infections is each infected student will create just less than one further new case. For example, for a prevalence rate of 15% we would expect approximately  $1000 \times 0.15 = 150$  secondary cases, which is within the error bounds for the simulated value of 141 secondary cases. Moreover, dividing each of the mean  $T$  values by the corresponding prevalences,  $I$ , and student population,  $N$ , provides a consistent probability of secondary infection (given an infected student) of  $H \times S \approx 0.94$  (95% CI 0.69 – 1.19). Since  $p \approx 1$  this justifies our observation. Finally, considering equation (1) and using the mean values from the literature we find that  $T \approx 0.91N$ , which again justifies the observation that each student infects just less than one other household member, on average.

Until now, we have quoted the number of new cases per 1,000 students. However, using the average value of secondary infections to be 0.94 times the number of infected students, we can expand this out to the entire student population of Wales. Specifically, it is estimated that there are 99,900 households in Wales that contain at least one student in Higher Education. Although we do not readily have access to data on how many of these students live at home, versus those that live away and will return home for the holidays we can take the 99,900 as an upper bound. Thus,

Student infection rate, $I$ (%)	Average number of new (secondary) cases, $T$ (per 1000 students) with 95% confidence interval
0.5	4.7 (0 - 9.8)
1.5	14.1 (5.3 - 22.3)
5	47.0 (31.1 - 62.9)
10	94.1 (71.9 - 116.5)
15	141.0 (114.1 - 167.9)

Table 3: Mean average number of cases expected for varying levels of virus prevalence. Results based on  $10^4$  repeated simulations of 1000 students.

from our results we would expect approximately 4,670 (at a  $I = 5\%$  infection level), or 1,409 (at a  $I = 1.5\%$  infection level), new cases stemming from secondary infections from students returning home. Extending this to the entire UK, where there is the potential for movement of over 1 million students for the winter vacation, even at a modest 1% infection level (meaning 10 in 1,000 students are infected, perhaps many of them without symptoms at the time of travel) that would equate to 9,400 new secondary household cases across the entire country.

One also has to be mindful of household compositions and thus not just consider the total number of secondary cases, but who the students live with. From Welsh data (Welsh Government 2020) we know that 52.9% of all student households contain at least one other person with a diagnosed long-term illness. Such people are at greater risk of hospitalisation. Thus, for example from the additional 4,670 cases calculated above (at a 5% infection rate), 2,470 of these cases would be in those with a diagnosed long-term illness.

## 5 Discussion

Universities cannot influence the makeup of a household. Equally, their influence on secondary infection rates is minimal. However, one potential suggestion is that universities could start their own media campaigns aimed at students' households. For example they could send leaflets to a student's registered home address that explains the risks that the household will face and steps that they can do to minimise their risk.

The only term within a university's power to practically control is the prevalence of the virus in the student body,  $I$ . Critically, with the insight that every infected student that is allowed to leave is expected to infect at least one other person, it is imperative that plans are made to reduce  $I$  to as sustainably low as possible, as quickly as possible.

### 5.1 Limitations

Critically, this work only accounts for transmissions and subsequent cases that might occur within a student's household. We do not consider transmission to the wider home communities in which the students live. Further, we do not include the journey home, which may give rise to a larger number of cases, particularly if public transport is taken. Thus, our numbers are a lower bound on the likely impact of transmissions and new cases. Furthermore, this work does not account for secondary cases, whereby students are infected by other members of their household.

The accuracy of our data must be considered. Namely, the data we have used is either based on Cardiff University asymptomatic testing, Welsh Government analysis, or “FF100” (first few hundred) cases UK data. Thus, even if we can be confident in providing predictions for Welsh universities we urge the reader to rerun our simulations with their own data, which can be achieved using the Matlab code, or online app. The code and app are quick to run with a focus on accessibility so that a user can rapidly change the input probabilities to suit their data.

## 5.2 Impact

Our work has already helped to influence government policy in relation to the firebreak (lock-down) in Wales during the period 23 October to 8 November 2020. Our model was used to assess the potential impact of individual, collective or institutional decisions for students to return from their halls of residence to their permanent home addresses. The analysis revealed that many of the students that would be affected live in multi-generational households and as a consequence this type of migration would expose older people living at the students’ home address to a higher level of risk of infection, ill-health and possible hospitalisation. Our analysis was considered by the Welsh Government’s Technical Advisory Group (TAG) and Higher Education Task and Finish Group dealing with Covid-19. On the basis of a consideration of the results of this work, Huw Morris (Director, Skills, Higher Education and Lifelong Learning, Welsh Government) confirmed that it was decided that there was evidence that encouraging students to return to their permanent home address from university residences would create greater risks than encouraging students to remain located close to their university of study.

Welsh Government also used our work for helping to evolve policy on the forthcoming winter vacation. Although the indicative levels of secondary infections are potentially very large, multiple strategies can be adopted to help reduce the number of students taking Covid-19 home. These include strongly advising students not to mix in the days leading up to departure, implementing staggered departure times and facilitating mass testing of students before they head home. Welsh universities have now been asked by Education Minister Kirsty Williams, to where possible put teaching online from December 8 to allow time for students to be tested prior to departure. This decision and the reasoning behind it, including the determined numbers of secondary cases from our findings, was communicated to colleagues in the English Government, Scottish Government and Northern Ireland Executive and informed the wider development of policy in this area across the UK.

## 5.3 Future work

Currently our equation can be used to produce the number of secondary infections that arise from students returning home. However, the equation elements can easily be redefined to provide estimates for secondary cases initially arising from students returning to university after the holidays. In this case,  $I$  would become the prevalence in the student’s community. This could either be taken as a UK-wide estimate, or a local estimate, depending on the specificity of data available.  $N$  can once again be varied, or set to 1,000, depending on the case of interest.

The household size,  $H$ , would have to be redefined. To what would be a difficult choice, but as one potential candidate there is the idea of “kitchen group” in Cardiff University’s halls of residence. Specifically, student rooms are spaced such that each kitchen is expected to be used by around 6 students. This would need to be combined with data on term time household sizes for those students who do not live in university accommodation.

Finally, defining the secondary probability distribution,  $S$ , is dependent of defining  $H$  appropriately. Alternatively, mathematical models, such as those discussed in Section 1, could be used to supplement our knowledge of  $S$  in the case that we do not have the appropriate statistics. Namely,

assuming some general, simple dynamics of student interactions, we could use ordinary differential equations to generate an effective reproduction number (Delamater et al. 2019, Ferguson et al. 2020).

Critically, it should be noted that altering the equation in this way would only provide an estimate for the initial number of secondary cases. Once an infection takes hold dynamical systems modelling would be needed to account for transmission to tertiary cases and beyond.

## 6 Strategies

It is our hope that these findings inform the discussion on allowing students to travel back home safely over the Christmas period. Specifically, we now provide a range of suggested strategies and present their positive and negative aspects.

If infection rates are very low then secondary infection rates will also be low, thus, a university could attempt not to put any leaving restrictions in place. However, this does depend on predicting what infection rates may be in December, which is difficult. Equally, this strategy is not robust as it only requires a sudden outbreak in a hall of residence, for example, for the strategy to fail, resulting in many students having to self-isolate and thus not return home for Christmas. Or, more problematically, the students decide to return home in spite of the sudden outbreak and risk high numbers of secondary transmissions. Moreover even if prevalence rates are low, any asymptomatic student returning home is a threat to any vulnerable household occupants. Thus, if no student facing strategy is to be implemented we strongly advise key messaging to households, which might help raise awareness of risks of secondary household transmissions, especially for those living with vulnerable people.

Such a non-strategy on the part of a university could be backed up by home testing. Namely, on arrival back home, the student would have a test and a further followup test 5 days later. In the mean-time a student would need to self-isolate in their bedroom, at home, until the second test result is known. This would require shifting the responsibility onto the students, with expectations that they would not socialise during this time. This strategy would lower the transmission risks when at home, but it does not combat the risk during the initial travel, which is particularly troublesome for international students. This home-testing strategy also raises serious capacity implications for massive volumes of testing in a relative short space of time, although staggered departure dates could help reduce this pressure.

In contrast to having no strategy a university could bring in a strict lockdown. Namely, for the last two weeks of term all students are to self-isolate. The benefit of such a response is that the students and their household can be confident in their safe return. However, there are obvious serious student compliance and well-being issues. Indeed, such a strategy might cause students to leave their university accommodation early, before the lockdown's implementation. The impact on a student's health, both physical and mental has to be considered in such an extreme situation, specifically because many students will have already been in periods of self-isolation Sahu (2020). There will also be additional concerns for students on certain practice-based courses, *e.g.* medical and dental degrees, as they will miss important placements and clinical practice.

The requirements of a strict lockdown could be softened if on campus testing could be achieved. Namely, if students can be tested immediately before returning home then any students prompting a positive result would have to self-isolate for 10-14 days (hence return home a little later than planned). Alternatively, students prompting a negative result should ensure they only go onto campus for face to face teaching sessions and avoid all end of term socialising. Finally, they should have a second test the day before they go home. If negative then they go home, whilst if positive they are required to self-isolate. Under this campus-testing strategy less time is spent in self-

isolation, perhaps prompting better compliance. Equally, the test results will reassure students and parents. However, compliance would still be an issue since we would require students to abstain from socialising during the run-up to Christmas. Equally, the ability to provide such a high volume of tests, as well as deliver results back in a timely manner, is a difficult logistic problem. However, as suggested with the home-testing strategy above, staggered departure dates could help reduce the pressure on the testing system.

Finally, if such campus-wide testing is not possible, but the prevalence is high enough that it is required then a university could focus its efforts on testing only those students who live with vulnerable people. This would dramatically reduce the amount of testing required, although it would cause new problems in estimating just how many tests would be needed. Equally, clear messaging to the students would be required to provide guidance on what counts as vulnerable.

The above suggestions are by no means a complete set of strategies. Neither are we suggesting that one plan of action will suit all universities due to their diversity of student body sizes and prevalence rates. Moreover, different universities maybe subject to different local governmental lockdown rules. The suggested strategies simply offer a broad set of guidelines that can be used and adapted, in conjunction with the data generated here, or by the reader generating their own data, to evidence a university's specific chosen policy.

## 7 Summary

Using a Monte-Carlo based approach and current data from the Covid-19 literature, we have been able to predict that each infected student that is allowed to return home is expected to produce (on average) just less than one further secondary infection. Our work has helped inform Welsh Government policy on the movement of university students in relation to both the firebreak (lockdown) in October/November 2020, and for the forthcoming vacation in December 2020. Although our results are heavily dependent on current available data, the method of result generation is robust and we offer the interested reader means by which they can rerun and adapt our simulations, thereby generating their own results based on localised parameters and thus for evaluating the associated risks of future movements of students to and from university settings.

## 8 Acknowledgements

JWM is supported by Knowledge Economy Skills Scholarships (KESS2), a pan-Wales higher-level skills initiative led by Bangor University on behalf of the Higher Education sector in Wales. It is part-funded by the Welsh Government's European Social Fund (ESF).

## A Monte-Carlo Matlab code

The following code has been run and tested on Matlab R2019a. The code simulates  $NS$  (in our case  $10^4$ ) stochastic samples of  $N$  students (in our case 1000). The code outputs the mean and 95% confidence intervals of the secondary infected population size.

The main inputs that can be modified by the user are: the probability distribution of household sizes, `Hhsizes`; the probabilities and standard deviations of secondary infections (depending on household size), `Mean_probs` and `Sd`, respectively; the number of simulated students,  $N$ ; the number of runs to average over,  $NS$  and the prevalence rates,  $I$ .

Currently, the code includes data for household sizes ranging from 2 to 6, as this was what was available from the cited literature. However, should the user want to try alternative data for household sizes and second infection rates they must ensure that the vectors, `Hhsizes`, `Mean_probs` and `Sd` are all the same length.

```

clear all
close all
clc

%% Input initialisation.
%% Defining the probability distributions.
% Percentage probability of household size (including student)
%      2      3      4      5      6+
Hhsizes=[22.1 27.6 27.7 10.1 4.3]/100;
% Note we ignore "single student" families, as they cannot infect anyone
% else.
% Data taken from Welsh Government analysis of the Annual Population Survey, 2019.
% https://gov.wales/annual-population-survey-july-2019-june-2020

% Rescale the probability boundaries to be 1.
Hhsizes=cumsum(Hhsizes/sum(Hhsizes));
Hhsizes=[0,Hhsizes];

% Mean probability (decimal) of infecting at least one other person in a
% household of size
%      2      3      4      5      6+
Mean_probs=[0.49 0.41 0.32 0.25 0.25];
% Standard deviation of probability distribution for a household of size
%      2      3      4      5      6+
Sd=[0.3617 0.3847 0.3535 0.2806 0.1255];
% Data from Tables 1 and 2 of "Transmission dynamics of Covid-19 in household and ...
% community
% settings in the United Kingdom" Lopez Bernal et al, 2020.
% https://doi.org/10.1101/2020.08.19.20177188

%% Simulation parameters
N=1000;           % Number of students
NS=1e4;          % Number of simulations
I=[15 10 5 1.5 0.5]/100; % Prevalence rates to be simulated

%% Monte-Carlo algorithm
for l=1:length(I)
    rate=I(l); % Run the code for each prevalence rate

    for j=1:NS % We average over NS simulations.

        % For each simulation, choose Nos uniformly randomly distributed numbers.
        % One for each simulated student. If a given random number is lower than
        % the prevalence then the simulated student is assumed to have
        % Covid-19. Thus, p1 is an indicator variable of whether each
        % student is infected.
        p1=rand(1,N)<rate;

        % Initialise p2 and p3.
        p2=[];
        p3=[];
    end
end

```

```

r=rand(1,N); % Generate Nos uniformly distributed random numbers to sample ...
    from the household distribution
for i=1:length(Hhsizes)-1
    p2(i,:)=i*(Hhsizes(i)<=r).*(r<Hhsizes(i+1)); % Sampled household sizes, ...
        providing number of susceptible people.
    p3(i,:)=min(ones(1,N),abs(Mean_probs(i)+Sd(i)*randn(1,N))); % Probability ...
        of secondary infection.
end

p2p3=sum(p2.*p3); % Average number of infected inhabitants assuming additional ...
    infected occupant.

Noi=p1.*p2p3; % Average number of secondary infections in each household.
Noit(j)=sum(Noi); % Total number of secondary infections
end

%% Calculate the mean and standard deviations of the total number of secondary ...
    infections
m(1)=mean(Noit);
sd(1)=std(Noit);
end

%% Output results as a table
Prevalence_in_percent=100*I';
Number_of_Secondary_infection=round(m,1)';
Standard_deviation=round(sd,1)';
Confidence_interval_lower=max(0,Number_of_Secondary_infection-1.96*Standard_deviation/sqrt(NS));
Confidence_interval_upper=Number_of_Secondary_infection+1.96*Standard_deviation/sqrt(NS);

table(Prevalence_in_percent,Number_of_Secondary_infection,...
    Confidence_interval_lower,Confidence_interval_upper)

```

## References

- Britton, T., Ball, F. & Trapman, P. (2020), 'A mathematical model reveals the influence of population heterogeneity on herd immunity to sars-cov-2', *Science* **369**(6505), 846–849.  
**URL:** <https://science.sciencemag.org/content/369/6505/846>
- Callow, K. A., Parry, H. F., Sergeant, M. & Tyrrell, D. A. J. (1990), 'The time course of the immune response to experimental coronavirus infection of man', *Epidem. Infect.* **105**(2), 435–446.
- Cardiff University (2020), 'COVID-19 case numbers in Cardiff University, accessed 4th November', accessed 4th November.  
**URL:** <https://www.cardiff.ac.uk/coronavirus/covid-19-case-numbers>
- Chahrour, M., Assi, S., Bejjani, M., Nasrallah, A. A., Salhab, H., Fares, M. & Khachfe, H. H. (2020), 'A bibliometric analysis of covid-19 research activity: A call for increased output', *Cureus* **12**(3).
- Christensen, H., Turner, K., Trickey, A., Booton, R. D., Hemani, G., Nixon, E., Relton, C., Danon, L., Hickman, M. & Pollock, E. B. (2020), 'Covid-19 transmission in a university setting: a rapid review of modelling studies', *medRxiv*.

- Currie, C. S. M., Fowler, J. W., Kotiadis, K., Monks, T., Onggo, B. S., R., D. A. & Tako, A. A. (2020), ‘How simulation modelling can help reduce the impact of covid-19’, *Journal of Simulation* pp. 1–15.
- Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T. & Jacobsen, K. H. (2019), ‘Complexity of the basic reproduction number ( $r_0$ )’, *Emerg. infect. dis.* **25**(1), 1.
- Ferguson, N. M., Laydon, D., N.-G. G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G. & Dighe, A. (2020), ‘Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand’, *Imperial College COVID-19 Response Team* .  
**URL:** <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf>
- Fishman, G. (2013), *Monte Carlo: Concepts, Algorithms, and Applications*, Springer Science & Business Media.
- Husky, M. M., Kovess-Masfety, V. & Swendsen, J. D. (2020), ‘Stress and anxiety among university students in france during covid-19 mandatory confinement’, *Compre. Psych.* **102**, 152191.  
**URL:** <http://www.sciencedirect.com/science/article/pii/S0010440X2030033X>
- Landau, D. P. & Binder, K. (2014), *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press.
- Latif, S., Usman, M., Manzoor, S., Iqbal, W., Qadir, J., Tyson, G., Castro, I., Razi, A., Boulos, M. N. K. & Weller, A. (2020), ‘Leveraging data science to combat covid-19: A comprehensive review’.
- Laxminarayan, R., Wahl, B., Dudala, S. R., Gopal, K., Mohan B, C., Neelima, S., Jawahar Reddy, K. S., Radhakrishnan, J. & Lewnard, J. A. (2020), ‘Epidemiology and transmission dynamics of covid-19 in two indian states’, *Science* **370**(6517), 691–697.  
**URL:** <https://science.sciencemag.org/content/370/6517/691>
- Lewis, N. M., Chu, V. T., Ye, D., Conners, E. E., Gharpure, R., Laws, R. L., Reses, H. E., Freeman, B. D., Fajans, M., Rabold, E. M., Dawson, P., Buono, S., Yin, S., Owusu, D., Wadhwa, A., Pomeroy, M., Yousaf, A., Pevzner, E., Njuguna, H., Battey, K. A., Tran, C. H., Fields, V. L., Salvatore, P., O’Hegarty, M., Vuong, J., Chancey, R., Gregory, C., Banks, M., Rispens, J. R., Dietrich, E., Marcenac, P., Matanock, A. M., Duca, L., Binder, A., Fox, G., Lester, S., Mills, L., Gerber, S. I., Watson, J., Schumacher, A., Pawloski, L., Thornburg, N. J., Hall, A. J., Kiphibane, T., Willardson, S., Christensen, K., Page, L., Bhattacharyya, S., Dasu, T., Christiansen, A., Pray, I. W., Westergaard, R. P., Dunn, A. C., Tate, J. E., Nabity, S. A. & Kirking, H. L. (2020), ‘Household Transmission of SARS-CoV-2 in the United States’, *Clin. Infect. Dis.* . ciaa1166.  
**URL:** <https://doi.org/10.1093/cid/ciaa1166>
- Lopez Bernal, J., Panagiotopoulos, N., Byers, C., Garcia Vilaplana, T., Boddington, N. L., Zhang, X., Charlett, A., Elgohari, S., Coughlan, L., Whillock, R., Logan, S., Bolt, H., Sinnathamby, M., Letley, L., MacDonald, P., Vivancos, R., Edeghere, O., Anderson, C., Paranthaman, K., Cottrell, S., McMenamain, J., Zambon, M., Dabrera, G., Ramsay, M. & Saliba, V. (2020), ‘Transmission dynamics of covid-19 in household and community settings in the united kingdom’, *medRxiv* .  
**URL:** <https://www.medrxiv.org/content/early/2020/08/22/2020.08.19.20177188>



- Mahmood, I., Arabnejad, H., Suleimenova, D., Sassoon, I., Marshan, A., Serrano-Rico, A., Louvieris, P., Anagnostou, A., Taylor, S. J. E., Bell, D. & Groen, D. (2020), ‘Facs: a geospatial agent-based simulator for analysing covid-19 spread and public health measures on local regions’, *J. Sim.* **0**(0), 1–19.  
**URL:** <https://doi.org/10.1080/17477778.2020.1800422>
- Manly, B. F. J. (2006), *Randomization, Bootstrap and Monte Carlo Methods in Biology*, Vol. 70, CRC press.
- ONS (2020), ‘Coronavirus (COVID-19) Infection Survey: characteristics of people testing positive for COVID-19 in England: October 2020’, *accessed 4th November* .  
**URL:** <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsinthecommunityinengland/october2020>
- Perez-Reche, F. & Strachan, N. (2020), ‘Estimating the number of covid-19 cases being introduced into uk higher education institutions during autumn 2020’, *medRxiv* .
- Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L. & Koole, M. (2020), ‘Online university teaching during and after the covid-19 crisis: Refocusing teacher presence and learning activity’, *Postdigi. Sci. Edu.* pp. 1–23.
- Ripperger, T. J., Uhrlaub, J. L., Watanabe, M., Wong, R., Castaneda, Y., Pizzato, H. A., Thompson, M. R., Bradshaw, C., Weinkauff, C. C., Bime, C., Erickson, H. L., Knox, K., Bixby, B., Parthasarathy, S., Chaudhary, S., Natt, B., Cristan, E., El Aini, T., Rischard, F., Campion, J., Chopra, M., Insel, M., Sam, A., Knepler, J. L., Capaldi, A. P., Spier, C. M., Dake, M. D., Edwards, T., Kaplan, M. E., Scott, S. J., Hypes, C., Mosier, J., Harris, D. T., LaFleur, B. J., Sprissler, R., Nikolich-Žugich, J. & Bhattacharya, D. (2020), ‘Orthogonal sars-cov-2 serological assays enable surveillance of low-prevalence communities and reveal durable humoral immunity’, *Immunity* .  
**URL:** <http://www.sciencedirect.com/science/article/pii/S1074761320304453>
- Rosenberg, E. S., Dufort, E. M., Blog, D. S., Hall, E. W., Hoefler, D., Backenson, B. P., Muse, A. T., Kirkwood, J. N., St. George, K., Holtgrave, D. R., Hutton, B. J. & Zucker, H. A. (2020), ‘COVID-19 Testing, Epidemic Features, Hospital Outcomes, and Household Prevalence, New York State—March 2020’, *Clin. Infect. Dis.* **71**(8), 1953–1959.  
**URL:** <https://doi.org/10.1093/cid/ciaa549>
- Rubinstein, R. Y. & Kroese, D. P. (2016), *Simulation and the Monte Carlo Method*, John Wiley & Sons.
- Sahu, P. (2020), ‘Closure of universities due to coronavirus disease 2019 (covid-19): impact on education and mental health of students and academic staff’, *Cureus* **12**(4).
- Seow, J., Graham, C., Merrick, B., Acors, S., Pickering, S., Steel, K. J. A. and Hemmings, O., O’Byrne, A., Kouphou, N. & Galao, R. P. (2020), ‘Longitudinal observation and decline of neutralizing antibody responses in the three months following sars-cov-2 infection in humans’, *Nature Microbiol.* pp. 1–10.
- Tan, W., Lu, Y., Zhang, J., Wang, J., Dan, Y., Tan, Z., He, X., Qian, C., Sun, Q. & Hu, Q. (2020), ‘Viral kinetics and antibody responses in patients with covid-19’, *medRxiv* .

- Teran, R. A., Ghinai, I., Gretsches, S., Cable, T., Black, S. R., Green, S. J., Perez, O., Chlipala, G. E., Maienschein-Cline, M. & Kunstman, K. J. (2020), ‘Covid-19 outbreak among a university’s men’s and women’s soccer teams—chicago, illinois, july–august 2020’, *69*, 1591–1594.
- Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. (2020), ‘A novel coronavirus outbreak of global health concern’, *The Lancet* **395**(10223), 470–473.
- Welsh Government (2020), ‘Annual Population Survey: July 2019 to June 2020, accessed 4th November’, *accessed 4th November* .  
**URL:** <https://gov.wales/annual-population-survey-july-2019-june-2020>
- WHO (2020), ‘WHO Statement – Older people are at highest risk from COVID-19, but all must act to prevent community spread’, *accessed 4th November* .  
**URL:** <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/statements/statement-older-people-are-at-highest-risk-from-covid-19-but-all-must-act-to-prevent-community-spread>
- Wu, J., Huang, Y., Tu, C., Bi, C., Chen, Z., Luo, L., Huang, M., Chen, M., Tan, C., Wang, Z., Wang, K., Liang, Y., Huang, J., Zheng, X. & Liu, J. (2020), ‘Household transmission of sars-cov-2, zhuhai, china, 2020’, *Clin. Infect. Dis.* . ciaa557.  
**URL:** <https://doi.org/10.1093/cid/ciaa557>
- Yamey, G. & Walensky, R. P. (2020), ‘Covid-19: re-opening universities is high risk’, *Br. Med. J.* **370**.  
**URL:** <https://www.bmj.com/content/370/bmj.m3365>
- Zhigljavsky, A., Whitaker, R., Fesenko, I., Kremnizer, K., Noonan, J., Harper, P., Gillard, J., Woolley, T., Gartner, D. & Grimsley, J. (2020), ‘Generic probabilistic modelling and non-homogeneity issues for the uk epidemic of covid-19’, *arXiv preprint arXiv:2004.01991* .