Mathematical modelling the emergence of viruses in naïve populations

Emma Louise Fairbanks

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

October 2021

Supervisory team: Professor Janet Daly, Professor Michael Tildesley, Doctor Marnie Brennan and Doctor Julia Kydd

School of Veterinary Medicine and Science
University of Nottingham

Abstract

Viruses pose a global threat to both humans and other animals. The severity of this impact is often amplified in populations with no prior immunity to the virus. Epidemics in naïve populations may occur due to novel virus emergence or the emergence of a known virus in a new geographical region. With climate change and globalisation, the frequency of the epidemics is likely to increase. Vector-borne viruses, spread by insects, have been increasingly observed in geographical regions where it was once believed that climatic conditions would not support the spread of these viruses. Mathematical models are a useful tool for policy makers to refer to while attempting to control and prevent epidemics. First, we focus on the *Culicoides*-borne virus African horse sickness virus (AHSV), using a systematic search to further the understanding of the dynamics of infection in naïve equines and parameterising a spatio-temporal model for the emergence of AHSV in naïve equids in Morocco 1998. Secondly, we apply mathematical and statistical models to study severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in universities.

Acknowledgements

First and foremost I am extremely grateful to my supervisors, Prof. Janet Daly and Prof. Mike Tildesley for their invaluable advice, continuous support, and patience during my PhD study. I would also like to thank the Biotechnology and Biological Sciences Research Council (BBSRC) for the funding opportunity to undertake my studies as well as the Nottingham BBSRC doctoral training programme team for providing support and opportunities to improve as a researcher throughout my PhD. My gratitude extends to Dr. Kirsty Bolton and the Issac newton SARS-CoV-2 higher education working group, from whom I have learnt so much. Finally, I would like to thank my family and friends for all their tremendous understanding and encouragement.

Contents

CHAPTER 1

Introduction

1.1 Overview

The aim of this project was initially to improve knowledge of the potential impact of African horse sickness (AHS) in the UK by building and analysing mathematical models. AHS, caused by African horse sickness virus (AHSV), is a vector-borne disease transmitted between equids and *Culicoides* spp.. As there have not been any previous outbreaks of AHSV in the UK, data are limited. Therefore, data from AHSV outbreaks in other regions, other vector-borne viruses and other diseases that affect equids have been considered. There is also the potential for this advice to extend beyond the UK to provide guidance across all regions and methods may also be applied to other vector-borne diseases.

Further knowledge on the behaviour and life cycles of European *Culicoides* as well as the effects of temperature on these would improve predictions. This information could also be used to aid in future forecasting as climate change proceeds. Future work was intended to include ecological modelling of UK *Culicoides* in order to inform models. Training was completed on *Culicoides* species identification and parity assessments at The Pirbright Institute. Pilot work trapping was conducted to establish the techniques for trapping and identifying *Culicoides* on equine premises (Appendix A). However, this work could not be concluded due to social-restrictions caused by the 2020 SARS-CoV-2 pandemic.

The global spread of SARS-CoV-2 resulted in the closure of workplaces, pubs and restaurants, restricted leisure activities and impacted the education sector. The first UK nationwide lockdown in March 2020 saw the closure of higher education establishments, such as universities, to most in-person activities. Universities are large communities of highly connected students and staff that distinguishes them from other types of workplaces. Despite mitigation strategies to reduce transmission risk, many UK universities experienced outbreaks of SARS-CoV-2 at the beginning of the 2020/2021 academic year. We used mathematical modelling and statistical techniques to predict the effect of control measures to reduce the risk of outbreaks within universities, as well as determine activities/behaviours that increase the risk of SARS-CoV-2 infection amongst university staff and students.

Mathematical and statistical methods that can quickly determine disease parameters in order to effectively react with control measures are increasingly being used by policy

makers. However, these methods, while they are often effective, have an uncertainty error that needs to be considered when they are used in decision making processes.

Outline of the work in this thesis

The main body of the thesis is presented as five research papers. The first three research chapters focus on infectious disease of equids. Chapter 2 is review paper that outlines the main pathogens that cause encephalitic disease in horses. The paper highlights management strategies such as surveillance and interventions to minimise transmission. Chapter 3 consists of a literature search to update knowledge on the key infectious disease parameters for models of AHS and uses them to parameterise an ordinary differential equation model considering host and vector interactions during an AHS outbreak. Chapter 4 examines a data set for the emergence of AHS in Morocco in 1989 and fits a spatial-temporal model of transmission between equid premises.

The final two research chapters focus on mathematical models of the spread of SARS-CoV-2 in university settings. Chapter 5 considers observational data from the first term of the 2020/2021 academic year and uses this to model control measures that were under consideration for the full return of UK higher education students in January 2021. Chapter 6 analyses survey results from individuals following an asymptotic PCR test for SARS-CoV-2 and uses Bayesian statistical analysis to associate the risk of contracting SARS-CoV-2 with different work and social activities. The impact of protective measures, such as mask wearing and social distancing, are also analysed.

In the remainder of the introduction we include general background material that was undertaken during the project, some of which is not directly referenced in the main body of the thesis.

1.2 African horse sickness virus (AHSV)

African horse sickness virus (AHSV) is a double stranded RNA virus of the genus *Orbivirus* of the family *Reoviridae*. There are nine immunologically distinct serotypes of AHSV, these are determined by the specificity of neutralising antibodies Fields et al. [2013]. Immunity from one of these serotypes does not protect against other serotypes and cross-infection is possible. The World Organisation for Animal Health (OIE) lists AHSV among its notifiable diseases; requiring member countries to monitor their disease status to avoid the spread of the virus [Hemida et al., 2017].

AHSV is endemic to sub-Saharan Africa with hosts including horses, mules, donkeys and zebra [Mellor and Hamblin, 2004]. Zebra, the only equids native to South Africa, are believed to be the reservoir host although large outbreaks have occurred in other regions were zebras are not present. The virus is transmitted between hosts by the vector *Culicoides* (Diptera: Ceratopogonidae). In Africa *C. imicola* is usually considered to be the most important vector species. However, in regions with different climates alternative species of *Culicoides* are likely to transmit the virus between hosts.

There are three forms of AHS; acute, subacute and subclinical [Carpenter et al., 2017]. If infected with the acute respiratory form, the host usually dies from severe hypoxia,

congestive heart failure, or a combination of both. In naïve populations of horses, mortality may exceed 90% [Castillo-Olivares, 2021]. The subacute cardiac form of AHS, with mortality rate around 50%, causes fever followed by oedema of the supraorbital fossae. A mixed pulmonary and cardiac form is most commonly seen in most outbreaks, which causes mortality of around 80% of susceptible equids. The subclinical form, often referred to as African horse sickness fever, is common in zebra, African donkeys and partially immune horses [Lu et al., 2020].

Due to the importance of horses in human history and the lethality of AHS, records as early as the fourteenth century identify this disease [Henning, 1949, Carpenter et al., 2017]. The first official recorded outbreak of AHS was in 1719 when 1,700 equids died. Outbreaks of AHS caused great strain on both the military and civilians, with an epidemic in Cape of Good Hope in 1854–55 causing the loss of 70,000 horses (approximately 40% of the local horse population) [Bayley, 1856, Carpenter et al., 2017]. AHSV was discovered to be a virus in 1900 [Fields et al., 2013]. In 1959–1961, a major epidemic spread through the Near East and Arabia as far as Pakistan and India, resulting in the death of an estimated 300,000 equids [Anwar and Qureshi, 1973, Howell, 1960, Carpenter et al., 2017]. In July 1987, AHSV-4 was the first serotype apart from AHSV-9 reported outside Africa, in central Spain. When lower temperatures began to restrict the ability of AHSV to spread in October it was hoped that this would eradicate AHSV from the region, however the virus successfully overwintered [Lubroth, 1988, Thompson et al., 2012]. In 1989, this epidemic spread to Portugal and Morocco. All three countries eventually eradicated the virus due to mass vaccination policies, strict movement controls and culling of equids [Portas et al., 1999, Rodriguez et al., 1992, Mellor, 1993, Baylis et al., 1997]. More recent outbreaks include AHSV-2 in Nigeria and Senegal and AHSV-7 in Senegal in 2007 [Diouf et al., 2013], AHSV-2, AHSV-4, AHSV-6, AHSV-8 and AHSV-9 in Ethiopia [Aklilu et al., 2014] in 2007-2010, and AHSV-1 in Thailand and Malaysia in 2020 [Castillo-Olivares, 2021, Lu et al., 2020].

1.2.1 Other *Culicoides*-borne virus

Due to climate change and increased globalisation, risk of *Culicoides*-borne virus outbreaks is increasing in some geographical regions such that these viruses now pose a threat to the UK [Li et al., 2021], with outbreaks of bluetongue virus (BTV) and Schmallenberg virus (SBV) having already occurred [Elbers et al., 2015, Wittmann and Baylis, 2000].

Bluetongue virus (BTV) Bluetongue virus is an *Orbivirus*, closely related to AHSV, also transmitted between ruminant hosts by *Culicoides*. There are twenty-nine serotypes of BTV [Yang et al., 2021] which causes clinical signs such as fever, swelling of the face, lips and tongue causing breathing difficulties if the tongue swells, reddening of the mucosal membranes, sores on the nose, gum and dental pads and lameness, and infection can be fatal [The Pirbright Institute, 2019]. BTV is teratogenic, sometimes causing deformities or abortion of calves and lambs by infection in utero.

Also an OIE notifiable disease, Bluetongue (BT) was discovered to be a virus shortly after AHSV in 1905 [Fields et al., 2013]. The first outbreak of BT outside Africa was in Cyprus in 1924, followed by another outbreak 1943–1944 with fatality among sheep

of approximately 70% [Polydorou, 1978]. With outbreaks occurring in the Middle East, Southeast Asia, Southern Europe and the United States in the 1940s and 1950s, BT came to be described as an emerging disease [Roy, 2008]. In 2006, cases of the virus where confirmed further north in Europe than previously observed [Toussaint et al., 2006]. Despite import controls and post import checks of ruminants to minimise the risk of importing an infected animal, the first UK case of BTV was confirmed on 22^{nd} September 2006 in Suffolk. By the 28^{th} September, the virus was confirmed to be circulating between ruminants and *Culicoides*. Mathematical modelling and weather observations suggested that this emergence in the UK was likely due to *Culicoides* been blown across the English Channel [Landeg, 2007].

BTV was estimated to have had a global consequence of three billion US dollars in 1989 [Rushton and Lyons, 2015]. Between 2002 and 2011 the EU funded 92 million Euros towards control and eradication of BTV. Between 2006 and 2008 in Germany, estimates for the impact of the BTV-8 epidemic ranged between 157 and million Euros (mean 180 million Euros). Only 27% of this was due to direct costs such as production losses, loss of trade, animal deaths, and veterinary treatment. The remaining 73% of indirect costs were mostly for vaccination (49%) but also included insecticide treatment, diagnostic testing, monitoring and surveillance and administration [Gethmann et al., 2020].

Schmallenberg virus (SBV) Schmallenberg virus (SBV) is spread by the same insects as AHSV and BTV, however it is in a different genus; *Orthobunyavirus* of the *Bunyaviridae* family. It is a negative-sense single-stranded RNA virus with a segmented genome. SBV emerged as a novel virus in 2011; close to the German/Dutch border. During late summer and early autumn there were reports of fever, decreased milk production, and diarrhoea for a few days in German and Dutch dairy cattle. In late autumn 2011, the first malformed lambs, which had been transplacentally infected with SBV, were born. This was followed by the birth of deformed calves starting at the beginning of 2012 [Wernike et al., 2014]. SBV spread very rapidly over large parts of Western Europe, with confirmed cases reported from 27 European countries by September 2013 at the end epidemic. Vaccines were used as an intervention in 2013, however sales quickly declined after the epidemic because it was not cost-effective and no DIVA assay was available [Wernike and Beer, 2020a]. In 2016, SBV re-emerged the UK, Ireland and Belgium along with other non-European countries [Collins et al., 2019, Wernike et al., 2014]. SBV continued to circulate in Ireland and the UK in 2017 and has since been reported in other European countries [Wernike and Beer, 2020b].

1.2.2 Equine encephalitic viruses

Viral encephalitis is one of the most common infections of the central nervous system (CNS) in horses worldwide [Barba et al., 2019]. Clinical signs can include mild fever, dullness, sleepiness, listlessness, ataxia, inability to rise, trembling, skin twitching, difficulty in urination and defecation, facial paralysis, blindness, seizures, coma, and other non-neurological signs [Long, 2014, Kumar et al., 2018, Barba et al., 2019].

Members of several different virus families cause equine viral encephalitis, the majority of which are arthropod-borne viruses (arboviruses) with zoonotic potential. The clinical signs caused are rarely pathognomonic; therefore, a clinical diagnosis is usually

presumptive according to the geographical region. However, recent decades have seen expansion of the geographical range and emergence in new regions of numerous viral diseases [Barba et al., 2019]. In Chapter 2 we present an overview of the prevalence and distribution of the main viral causes of equine encephalitis and discusses their impact and potential approaches to limit their spread.

1.3 Severe acute respiratory syndrome coronavirus 2

In December 2019, a cluster of patients with pneumonia of unknown cause was identified in Wuhan, China. This novel coronavirus, SARS-CoV-2, of the genus *Betacoronavirus* in the family *Coronaviridae* is the seventh member of this family able to infect humans including MERS-CoV and SARS-CoV) [Zhu et al., 2020]. The first case of SARS-CoV-2 in the UK was identified on 30 January 2020. The number of confirmed cases grew to 11,080 before a lockdown began on 24 March [Dropkin, 2020]. Lockdown measures included the closure of higher education institutes; including universities.

Prior to the return of students to campuses, preliminary modelling studies flagged universities as settings of potential high risk for SARS-CoV-2 transmission [Hill et al., 2021, Brooks-Pollock et al., 2021, Task and Finish Group on Higher Education/Further Education, 2020]. At the beginning of the 2020/2021 academic year, many universities experienced outbreaks of SARS-CoV-2 [UCU, 2020]. Chapter 5 presents work on SARS-CoV-2 transmission in UK higher education settings using multiple approaches to assess the extent of university outbreaks, how much those outbreaks may have led to spillover in the community, and the expected effects of control measures. This work was completed collaboratively with the Higher Education working group at the Isaac Newton Institute. In Chapter 6, we identify setting-specific contact measures and protective behaviours associated with risk of PCR-confirmed asymptomatic SARS-CoV-2 infection in a sample of university staff and students.

1.4 Epidemiological Models

Mathematical models can be a useful tool in predicting the behaviour of disease outbreaks; allowing us to understand the factors that drive epidemics [Chubb and Jacobsen, 2010]. They can be used to predict parameters we are unable to attain from laboratory studies. Daniel Bernoulli was the first scientist to mathematically model the spread of disease in 1766 in order to defend inoculating against smallpox [Hethcote, 2000]. Building models requires a trade-off between accuracy (reproducing observed and predicting future dynamics), flexibility (adapting the model for different situations/diseases) and transparency (ability of others to understand the model). A good model will balance all of these factors. However, these models are a simplification of reality and random events within disease transmission mean perfect prediction is not possible. All models must make assumptions and have limitations, and this must be recognised. Certain rare events, such as the introduction or emergence of variants and secondary introductions, and human behaviours can significantly impact the epidemiology and persistence of these viruses. For example, more recently discovered serotypes of BTV support vector-free transmission through contact and *in utero* Pullinger et al. [2016], this may impact the ability of the virus to over-winter and persist in the field. Models should be as simple as possible to describe the dynamics

we are interested in, but no simpler. It is also important, where possible, that we can parameterise models from the data available. Generally, the more complex a model is the more difficult it is to parameterise. In Appendix B we demonstrate a method for accessing the identifiability of models and Appendix C explores the influence of these parameters on model outputs.

Epidemiological models can be used to investigate the likelihood of virus transmission within the UK throughout the year to improve the effectiveness of surveillance and to minimise future risk of disease outbreaks. They can provide epidemiological forecasts, assess current and future outbreak risk and quantitative assessments of the effectiveness of intervention strategies. Often epidemiologists are interested in the basic reproduction number, R_0, of a disease. R_0 is the expected number of new infectious cases caused by one individual during their infectious period [Dietz, 1993]. Therefore, if R_0 is larger than one, the number of cases would increase and therefore an outbreak would be expected to occur, whereas if R_0 is less than one we would expect the disease to die out. In Appendix D we use mathematical models to explore the emergence of AHSV in a naíve population (Morocco, 1989) and compare the vector capacity and R_0 of AHSV, BTV and SBV in the UK.

1.4.1 Temporal models

The host-to-host transmission of a disease within a population can be modelled using a deterministic SIR model. Here S, I and R represent the number of individuals susceptible, infectious and recovered from the disease within the population, respectively. Individuals are considered susceptible if they are not infected or immune to the disease, therefore they could still become infectious. If these individuals become infected they move to the infectious stage, and then to the recovered stage when they are no longer infectious. We then consider the rates at which hosts transition between stages (infection, recovery, etc.). Here infection would be represented by the transition

$$S + I \longrightarrow 2I,$$

and recovery would be represented by the transition

$$I \longrightarrow R.$$

The rate at which these occur is derived using the law of mass action. The law of mass action was first proposed by Cato Maximilian Guldberg and Peter Waage in 1864 [Guldberg and Waage, 1986]. Today this law is well known and has a wide variety of applications across many areas of science. In the case of infection, the law would state that new infectious cases arise at a rate proportional to the number of susceptible and infectious individuals in the population and recovery occurs at a rate proportional to the number of infectious individuals. The constant of proportionality is called the rate constant. The first modern application of this model was by George MacDonald who modelled the spread of malaria [Macdonald, 1957]. Figure 1.1 shows how individuals move through the stages.

The rate of transition between stages depend on the epidemiology of the virus. For example, the rate of recovery depends on the infectious period, and the rate of transmission depends on the infectiousness of the virus. There are methods of reducing

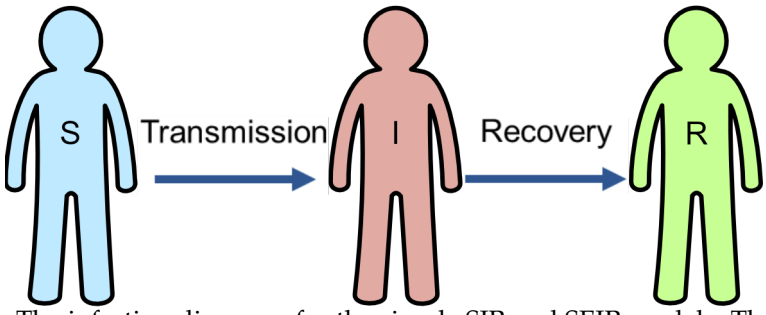

Figure 1.1: The infection diagrams for the simple SIR and SEIR models. The stages of infection are given by the boxes. Transition between the stages is represented by arrows with their rates given above them.

these rates, for example, reducing the infectious period of infectious hosts by isolation and reducing the transmission rate by reducing host contacts.

These models can be adapted to include more complex dynamics, e.g. hosts not enquiring immunity from infection, latently infected hosts or super-spreader hosts, which have a larger viral road. When adapting these models to include vectors we also need to consider the dynamics of the vector populations as well as the interactions between vectors and hosts. In Section 1.6 we discuss climatic influences on the life-cycle of *Culicoides* and their ability to spread viruses. Chapter 3 uses data from a systematic literature search in order to better inform transition rates for a previously published ordinary differential equation model of AHSV within a premises.

Another popular method of disease modelling, considered in Chapter 5, is network graphs. Network graphs can also be used to model potential disease spread. Here nodes represent hosts and edges represent connections between hosts allowing for disease transmission. At each iteration, the neighbours of an infected node are assigned a probability of infection and infected nodes have a probability of recovery. These edges can be weighted allowing for differences in the probability of infection between two individuals depending on their contact rates. These methods are less popular when modelling *Culicoides*-borne viruses due to the large vector population increasing the computational demand for this algorithm.

1.4.2 Spatio-temporal models

Often we are interested in the spatial dynamics of disease transmission. Although there are other methods, including partial differential equations, kernel-based methods are becoming increasingly popular [González et al., 2016a]. Kernel-based models can be used to assess the risk of disease spread by considering the distance between infected and uninfected locations. These locations could be individual premises, small areas (for example post codes) or regions/countries. The distance kernel, $K(d_{ij})$, describes how the transmissibility of a virus varies in relation to the distance of an infected location (j) from a naive location (i). The distance between two locations d_{ij}, can be described by the Euclidean distance between them. For between premises models the latitude and longitude co-ordinates of the premises are used to calculate this distance and for larger-scale models the distance between the centroids of the areas are often used. Using these kernels, we calculate the force of infection, $\lambda_i(t)$, for location i on day t typically given

by

$$\lambda_i(t) = \beta\theta(t)N_h^{(i)}\sum_{j\neq i}K(d_{ij})N_h^{(j)}I_j(t), \tag{1.1}$$

where β is the transmission parameter and the variable $I_j(t)$ is 0 or 1 depending on whether location j is uninfected or infected, respectively, on day t. The parameter N_h represents the number of hosts at the locations for between premises models and some larger scale models, however sometimes the number of premises is used here. This model can be used to compare the susceptibility and infectivity of species, for example parameterising

$$\lambda_i(t) = \beta\theta(t)(N_{h,1}^{(i)} + mN_{h,2}^{(i)})\sum_{j\neq i}K(d_{ij})(N_{h,1}^{(j)} + nN_{h,2}^{(j)})I_j(t) \tag{1.2}$$

would give the infectivity and susceptibility, m and n respectively, of the host species 2 $(h, 2)$ compared to host species 1 $(h, 1)$.

When applying kernel-based approaches to agricultural animal disease, such as foot and mouth disease, detailed information on the location of farm premises and animal movements is available. Knowledge of the UK equine population demographic and movements are discussed in Section 1.5. In Chapter 4 we fit a kernel-based model to data from the 1989 emergence of AHSV in Morocco.

1.5 Equine location and movement

The location and transport of cattle and sheep is well documented, requiring a licence to keep or move these animals. However, there is no requirement for equine premises to be registered [Owers and Meldrum, 2013]. Equine movements within the UK could help aid the spread of an infectious disease during an epidemic. International movements could lead to an import of an equine infectious disease. The outbreaks of AHSV in Spain in 1987 and Thailand in 2020 were associated with importation of (infected) zebra from Africa [Grewar et al., 2021].

EU regulations and OIE guidelines state disease freedom can only be reinstated after a period of two vector free seasons without evidence of virus circulation. Controls on registered equines (and registered equine product) imports to the EU from any third country to the EU include pre-export tests, 40 day vector free quarantine before export and certification stating the animal's vaccination and test status [Department for Environment and Rural Affairs, 2016]. South African equine imports are approved from only a small area of the City of Cape Town dependent on surveillance of horses in the surrounding region. Equines may move from a AHS free area as long as they have been resident for at least 60 days.

1.5.1 The uncertainty of equine location in the UK

The UK's National Equine Database (NED) contained information on the equine population including: number of equids and distribution of their owners. In June 2010, NED had records of 1,383,304 equines with passports. Boden et al. [2012] found that of these only 78% had valid addresses (with equids being excluded for having 'no recorded address', 'no fixed abode', no valid matching postal code, having foreign passports

and/or foreign addresses). On mainland GB, 1,034,907 equids had owners with a valid postcode area [Boden et al., 2012]. Although there are likely to be a considerable number of equids over the age of 30 in the UK (life expectancy 25-30 years [Ensminger, 1990]). Boden et al. [2012] estimated the population of equines in mainland GB which are most likely to be alive (30 years old or less) and have valid postcodes as 842,653 equids [Boden et al., 2012]. This would correspond to a non-declaration of death of an animal of 19%. Robin et al. [2013] found during 2005–2010, that of 17,048 passports checked through local authority inspections, 9.1% were non-compliant, with 5.6% containing inaccurate information and 3.5% classified as missing. Of 1382 owner questionnaires, 27.5% were obsolete (11.7% being retained for deceased horses and 15.8% having incorrect ownership details) [Robin et al., 2013].

NED recorded the location of owners rather than equids. This is the main issue with using this database for spatial analysis of potential equine disease outbreaks. NED ascribed 7,432 equids (104 per 10 km^2) to London addresses, the highest in England. However, Boden et al. [2012] found that a dataset created using data from equine stakeholders (based on equine location rather than owner location) ascribed 1,749 equids to London (25 per $10km^2$), the lowest density of horses in England. In Wales, although Cardiff has the largest density of equines in both datasets, this density was also much lower in the stakeholder dataset (NED: 102 per $10km^2$, stakeholder: 72 per $10km^2$). Whereas in Scotland the equine density was highest in Kircaldy in both datasets, however the stakeholder dataset predicted that there where more equids in this area (NED: 23 per $10km^2$, stakeholder: 32 per $10km^2$) [Boden et al., 2012]. This demonstrates that owner location cannot always be viewed as a direct substitute for horse location. Robin et al. [2011] found that 61% of a sample of 1440 equids were kept at the same postcode as their owners. Overall, the mean $\pm$ standard error distance between equines and their owners addresses was 8.11 ± 1.1km, including kept at the same postcode as their owner (20.82 ± 2.84km excluding equines kept at the owner address). Only 2.5% of equines were kept over 50km, while 90% of equines where kept less than 10km from their owners [Robin et al., 2011]. Another survey study of Great British equine owners (n=3966) found that 93% kept their equines within 16km of their location, with 6% being kept between 16 and 80km and 1% being kept and more than 80km away from their location [Boden et al., 2013].

Robin et al. [2013] evaluated the spatial separation between horses and their owners and identified relationships between this spatial separation and land use using questionnaires (1010 samples). Analysis of the data showed that the distribution of the spatial separation between horses and their owners was well described by a power-law distribution, irrespective of the local values of built-up coverage. Heavy-tailed distributions allow large separations between horses and their owners while allowing most horses to be located close to their owners (assuming horse premises near the owner's location are preferable) [Lo Iacono et al., 2013, Robin et al., 2011, 2013]. Lo Iacono et al. [2013] estimated the distribution of equids in the UK generated by considering the urban coverage [Fuller et al., 2002] of the owners address to place equids in 'more likely' locations. It was shown that mapping owner's addresses as a proxy for horse location significantly underestimates the risk of disease transmission. Incorporating this into models would facilitate a more meaningful distribution derived from NED for modelling the risk of equine disease outbreaks in the UK.

For economic reasons, Defra controversially ceased their funding (around £200,000/annum) of NED in September 2012 [Owers and Meldrum, 2013].

The Lo Iacono et al. [2013] equine distribution estimates the number of equids per 5 *km* block of the UK. However, nothing is inferred about the number of equids on each premises. In some locations, equids are likely to be kept together, with multiple owners keeping equines on the same premises. It would also be expected that equids would be housed in areas with suitable grazing and therefore equine premises are likely to be clustered.

1.5.2 UK equine national and international movement

Boden et al. [2013] used an online questionnaire to obtain unique information linking owner and horse location to characteristics of horse movements within Great Britain (GB) and internationally. During the year proceeding the survey 59% of respondents (n=3522) travelled and returned home with their horse within a single day. Of the 42% of respondents that travelled with their equids for more than one day, 71% were away 1–7 days, 24% were away 8–30 days, 5% were away from the home premises for more than 30 days. Of 3541 respondents, 6.3% travelled internationally with their equines and/or imported horses from Belgium, Ireland, Germany, Spain and Poland. Owners who travelled internationally were three times more likely to have at least one horse with a foreign passport rather than a British-issued horse passport (unadjusted OR 3.4, 95% CI 2.5–4.5 p-value <0.001).

1.5.3 The uncertainty of and attitudes towards equine location/movement databases in non-UK countries

France The French equine database (SIRE) records information on equines, including contact details of the owners [dor]. It also records information on "equine premises", where equines are kept, because French regulations require keepers to notify the authorities of the opening and/or closing of such premises (Article R215-14 of the French Rural Code) [Farchati et al., 2021]. However, as with many countries databases, owners/keepers do not systematically comply with these regulations meaning some of the information in the database is incorrect. Farchati et al. [2021] carried out surveys of owners and keepers that had given an email address and agreed to contact; 20.1% of owners and 19.5% of keepers responded. The rate of non-declaration of an equine's death in the survey was 6.2%, therefore the number of equines in France is likely to be overestimated. Concerning the reference equid, 11% of respondents stated that they no longer owned this animal and 3.6% reported that they no longer owned any equids. Of the respondents, 64.7% had the same postcode in the survey and the SIRE database. The owner survey indicated that 52.6% hosted equids but were not registered as keepers. Overall, the under-declaration of equine deaths and premises would potentially restrict capability of surveillance and epidemiological investigation.

United States of America The US National Animal Identification System (NAIS) is an animal tracking database created to identify and record the movement of animals

in the event of an animal disease outbreak. A survey administered to American Association of Equine Practitioners (n=139) showed that 55.6% thought that NAIS would help efficiently stop the spread of a contagious animal disease, however only 19.4% were “very familiar" with the database [Vanderman et al., 2009]. Another survey was designed to evaluate equine event official demographics (n=115). Here 24% of show managers reported being in favor of NAIS, 18% were opposed and 58% were neutral or unsure, with 53% reporting that NAIS would increase paperwork at events [Swinker et al., 2009].

Denmark Hartig et al. [2013a] found that 68% of stakeholders (n=698) thought a national equine health database was important, including equine veterinarians, researchers, representatives from animal welfare organizations, horse owners and others. Most stakeholders wanted the database to focus on contributing to improved horse health and welfare rather than on performance or food safety. Requirements for such a database included an electronic, simple, and time-efficient data reporting system. Overall, it was found that there was a positive attitude to the establishment of a health database in Denmark [Hartig et al., 2013a,b].

1.6 *Culicoides*

There are forty-eight known *Culicoides* (Diptera: Ceratopogonidae) species in the UK [Boorman, 1986]. However, not all species are capable of transmitting viruses. In Africa the main vector for BTV and AHSV is *C. imicola*, however other species are capable of spreading the viruses [Foxi et al., 2016]. In the Americas *C. sonorensis* and *C. insignis* are the main vectors for BTV (depending on the region), and in Europe *C. obsoletus* [Gerry and Mullens, 2000].

Culicoides life-cycles include the physiological stages: the egg, four larval instars, the pupa and adulthood (Figure 1.2). The duration of these stages varies between species and is dependent on temperature. During the larval and pupal stages moisture is required, therefore eggs are commonly laid in semi-aquatic habitats [Carpenter et al., 2013]. In northern Europe *Culicoides* generally overwinter as fourth instar larvae [Blackwell et al., 1992, Hope, 2013]. Fully-grown *Culicoides* are approximately 1–3 mm long. Adult males only feed on nectar, whereas females also feed on blood. Females that are yet to lay eggs are described as nulliparous. Whereas those that have previously laid eggs are described as parous, identifiable by a change in pigmentation in their abdominal cuticles and tergites [Gerry and Mullens, 2000].

Relatively small changes in the expected vector lifespan have theoretically been shown to have an effect on transmission of pathogens [Gerry and Mullens, 2000]. For *Culicoides*-borne viruses, the midge ingests the virus during a blood meal; the virus must then spread within the insect, involving several rounds of infection and replication in different tissues before it reaches the salivary glands (or possibly the mouth parts), causing the midge to become infectious. The extrinsic incubation period (EIP) is the interval between the acquisition of an infectious agent by a vector and the vector being able to transmit the agent to susceptible hosts [Dictionary, 2007]. The rate at which

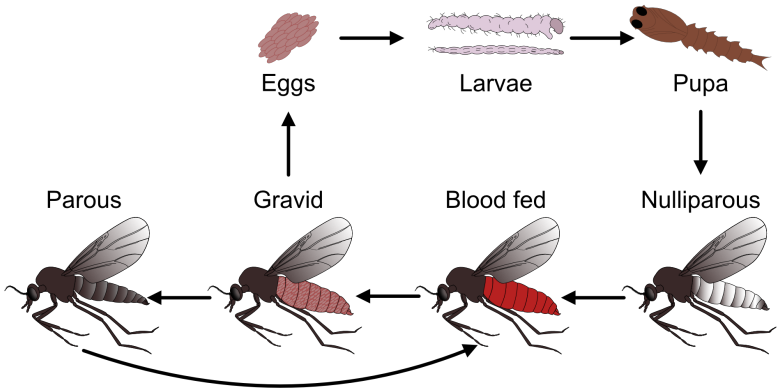

Figure 1.2: *Culicoides* life-cycle and gonotrophic cycle.

these viruses replicate is dependent on temperature, therefore as *Culicoides* are poikilothermic, in temperate regions this imposes geographical and seasonal limits on the spread of these viruses.

Various studies have tried to quantitatively assess how temperature and other climatic effects determine the ecology of *Culicoides* and epidemiology of the viruses they transmit. Here we give an overview of some of the the mathematical models used to explain these dynamics from a UK perspective.

1.6.1 Extrinsic incubation period

Cooler climates restrict virus replication, and therefore extend the EIR, reducing transmission. Carpenter et al. [2011] developed statistical methodology for estimating the relationship between temperature and the duration of the EIP [Carpenter et al., 2011]. This method was applied to both published and new data [Paweska et al., 2002, Wittmann et al., 2002, Carpenter et al., 2011]. *Culicoides* are fed infected blood in a laboratory and kept at a constant temperature, groups of *Culicoides* are then removed at certain time-points and their infectiousness measured. The model for the temperature dependence of the EIP completion rate was given by,

$$v(T) = max\big(0, \alpha(T - T_{min})\big), \tag{1.3}$$

where the rate at which the virus replicates is assumed to increase linearly at rate α above a threshold temperature, T_{min}. Once the cumulative sum of the daily values of v reaches 1 the *Culicoide* is infectious. In other words, $1/\alpha$ is the number of degree-days above the threshold temperature required for the vector to complete the EIP. Many of these studies used colony derived insects and may not be fully representative of field populations from different regions, or different species.

It was found that there can be differences in how well each *Culicoides* species can replicate each virus, and how well each virus replicates in each species. They also have a different probability of becoming infectious from being fed a blood meal. This is often equated to the probability of *Culicoides* becoming infectious when biting a host in models. It was found that the virus replication rate and threshold temperature were similar for a range of bluetongue virus (BTV) and African horse sickness isolates [Carpenter et al., 2011]. However, the probability of infection was higher for African horse sickness

virus (0.52) than any of the BTV strains (0.12–0.16) in *C. sonorensis*.

It was also been found that the percentage of infected *Culicoides* varies with level and duration of viraemia in the blood meal [Mertens et al., 1996], this varies over the course of infection, with different virus strains and serotypes of virus, and between individual hosts. The level and duration of viraemia can be reduced by vaccination helping to prevent onward transmission by individuals that still become infected. The level of viraemia may be higher for more virulent strains, which could also cause a higher rate of fatalities. Immune responses from other strains or serotypes may also affect level and duration of viraemia and death rates within the hosts, by either partial immunity or antibody-dependent enhancement. These factors may be particularly important in endemic areas.

1.6.2 Bite rate

The gonotrophic cycle is defined as the time required for host location, blood feeding, egg maturation and oviposition. This has been observed to be affected by temperature. The time between blood feeding and oviposition (O), for *C. sonorensis*, has been investigated for temperatures 13–34 °C and was found to follow the regression equation:

$$O(T) = -1.98 + 0.07217T + 2516.65T^{-2} \tag{1.4}$$

[Mullens and Holbrook, 1991]. Figure 1.3 shows this regression plotted for -5–35 , these temperatures would be more realistic in the European climate. This equation increases without bound as it approaches $T = 0$. We observe that while this curve may fit the data well from 13–34 °C it probably does not give an accurate estimate for lower temperatures. However, as *Culicoides* over-winter as larvae in European climates, this may not cause unusual model behaviours. The *Culicoides* used in the study were also kept at constant temperatures, which is not realistic for field *Culicoides* [Mullens and Holbrook, 1991]. Mullens et al. [2004] also attempted to fit these data to approximate the ovarian development rate (σ), defined as 1/time to oviposition. The model suggested was

$$\sigma(T) = max\big(0, 0.00019T(T - 3.6966)(41.8699 - T)^{1/2.7056}\big), \tag{1.5}$$

where T is the temperature. Although this model often predicts slightly shorter times to oviposition it still exhibits the same unstable behaviour for low temperatures. By assuming that each midge only bites once in order to complete its gonotrophic cycle we can equate the length of this cycle to the frequency that *Culicoides* bite.

1.6.3 Adult lifespan

As the length of the EIP decreases with temperature so does the expected lifespan of the vector, therefore higher temperatures may not always result in more disease transmission. MacDonald [1952] estimated the probability of an insect surviving the EIP as $p^{(1/v)}$, where p is the daily survival probability and $1/v$ is the EIP [MacDonald, 1952]. How long the insect will survive after completing the EIP was estimated as $p^n/-ln(p)$ [MacDonald, 1952].

Gerry and Mullens [2000] attempted to quantify how the expected lifespan of a midge varies according to temperature. Here the lifespan was estimated using the parity rate, P = number of parous females/total number of females, known as the mean parous rate

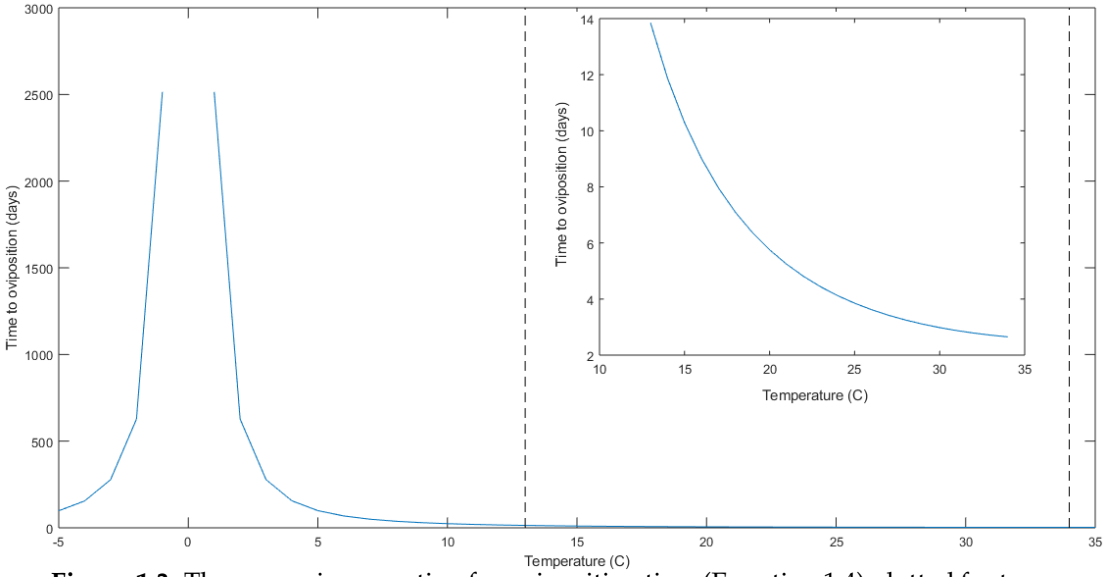

Figure 1.3: The regression equation for oviposition time (Equation 1.4) plotted for temperatures $-5 \leq T \leq 35$ °C. Black dotted lines represent the limits for which the regression was fitted to data. The regression is also plotted separately (smaller plot) within this range to highlight the values the original model predicted. Since this function is infinity at 0 values $-1 < T < 1$ are not plotted.

formula [MacDonald, 1952]. The probability of daily survival is estimated as $p = P^{1/u}$, where u is the length of the gonotropic cycle (calculated from Equation 1.4). Over the 3-year study, *C. sonorensis* were captured on a Californian (USA) dairy farm. The amplitude of abundance of females varied with year, mean monthly abundance was significantly correlated with mean monthly temperature, but host-seeking females were captured throughout the year. Overall, monthly parity was negatively correlated with mean monthly temperature and mean monthly female abundance. The length of the predicted gonotrophic cycle varied from 3–4 days in summer and 14 days in winter (Equation 1.4). The daily probability of survival varied from 0.53–0.97, and was higher in colder months. Using these data, the expected lifespan was calculated, decreasing exponentially with increasing temperature. This was as short as 2 days in summer and was generally longer than 10 days in winter. However, the region this dairy farm was in has a relatively high BTV prevalence. The $75km^2$ region containing 250,000 milking cows in which most of these dairies have wastewater ponds close to cattle provides an ideal breed site for *Culicoides*. This ideal *Culicoides* environment may however not be representative of many other habitats, including European farms; therefore this lifespan may be an over-estimate.

Overall, the study predicted the expected lifespan (eLS) to be:

$$eLS(T) = 111.84exp(-0.1547T), \tag{1.6}$$

where the mean monthly temperature was used (Figure 1.4a). The original study was based on mean monthly temperature, whereas the function is often used on daily temperatures (eg. [Gubbins et al., 2014]). We would expect daily temperatures to have larger extremes and Figure 1.4 shows that this equation exponentially increases as temperature decreases, however no data have been collected for lower temperatures to

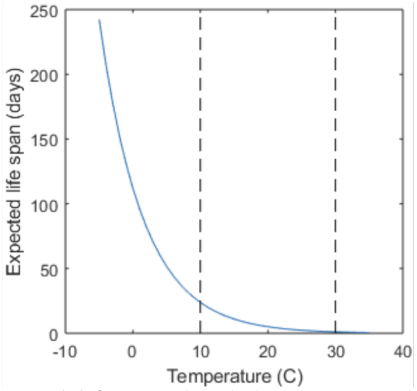

Figure 1.4: *Culicoides* expected lifespan (*eLS*) (Equation 1.6). The black dotted lines represent the region in which data fitted.

clarify this relationship. Mean temperatures in Europe are often outside this range. Also, the study was conducted on the species *C. sonorensis*, which may not be an accurate representation of all *Culicoides* species.

Birley and Boorman [1982] attempted to modify MacDonald's method to more accurately estimate survival rates. The number of parous females (Q) that bite at time t is equivalent to the number which that fed one cycle earlier multiplied by the number which survived that cycle, given as:

$$Q(t) = SM(t-u), \tag{1.7}$$

where S is the survival rate per cycle, u is the length of the gonotrophic period, and M is the total number of females. Thus the survival rate is given as

$$S = \frac{\sum Q(t)}{\sum M(t-u)}. \tag{1.8}$$

This model assumes: (i) there is no sampling bias between parous and nulliparous samples; (ii) a constant proportion of the population is sampled; (iii) all deaths are natural; (iv) only one blood meal is required for each female to complete the gonotrophic cycle; (v) changes in Q are only caused by the number of nulliparous females recruited; (vi) all females have a similar gonotrophic cycle length during the experimental period [O'Connell, 2002]. Equation 1.7 predicts that the time series $Q(t)$ and $M(t-u)$ are correlated [Birley and Boorman, 1982]. The gonotrophic period (u) was therefore estimated by deriving a cross-correlation index which maximises cross-correlation (the similarity as a function of displacement) [Birley and Boorman, 1982]. In reality the length of this cycle would be expected to vary between individuals. Using this method Birley and Boorman [1982] estimated that the mean gonotrophic period for *C. obsoletus* in southern England varied between 3.5 and 4.5 days between June and October. Mean daily average temperatures in the region of trapping would be excepted to vary between 11 and 18 °C this time of year [Met Office]. However, predictions from Equation 1.4 would predict the time between blood meals and ovipositioning to be 7–20 days at these temperatures. This difference could be due to a number of factors, including the species of *Culicoides* and laboratory *Culicoides* are often kept at a constant temperatures, which does not accurately reflect the climates field *Culicoides* are exposed to.

1.6.4 Seasonal abundance of adult *Culicoides*

Sanders et al. [2011] explored the seasonal variation in abundance of adult *Culicoides* and the influence of meteorological conditions. The season vector activity on day t, $\theta(t, M)$, was found to be,

$$log(m(t)) \propto b_0 + a_1 sin(\theta t) + b_1 cos(\theta t) + a_2 sin(2\theta t) + b_2 cos(2\theta t) + \sum_m c_m M_m + \sigma_{jk} + \epsilon_{jk}. \quad (1.9)$$

The trigonometric terms and b_0 describe the seasonality with periods of 12 and 6 months ($\theta = 2\pi/366$), the summation allows for the influence of meteorological variables (with parameters c_m), σ allows for over-dispersion in the data and ϵ allows for temporal autocorrelation between observations. The meteorological variables (M) at sunset included were temperature, mean wind-speed and whether it was raining [Sanders et al., 2011]. This model is often used without taking meteorological conditions into account. Here only b_0 and the trigonometric terms are used to represent vector abundance. This simplifies the model as fewer data are used and the computational cost is reduced, however this does not allow for variation between different regions. Although rainfall is an important factor in transmission of vector-borne diseases, it is very difficult to predict the time at which it will occur, especially for further into the future. Previous models have also found precipitation to have a weak effect on model outcomes [Brand and Keeling, 2017].

Turner et al. [2019] simplified the model proposed in Equation 1.9 to:

$$\hat{m}(t) = exp\{b_0 + p_1 sin(\theta(t - \phi_1)) + p_2 sin(2\theta(t - \phi_2)) + cT\}, \quad (1.10)$$

where T is the temperature and $\theta = 2\pi/366$. Here b_0 is a constant, the first trigonometric term represents the 6-month period and the second the 12-month period. The only meteorological condition considered is temperature; reducing the data required this function produces a seasonal pattern which combined with the bite rate (a) reasonably represented the pattern of trap catch data [Turner et al., 2019]. It is assumed that the number of trap catches represents the number of *Culicoides* per host per night and the proportion of *Culicoides* capable of spreading the virus is not considered.

Culicoides abundance has a strong impact on the spread of *Culicoides*-borne viruses. Gubbins et al. [2007] found that values of $R_0 > 1$ were associated with larger vector:host ratios. However, there are other factors to consider. For example, active *Culicoides* have been found at 7.5°C, however they cannot transmit a virus if the temperature is not high enough for replication [Nielsen, 1963, Carpenter et al., 2011].

1.7 Summary

"One model fits all" is not applicable to mathematical and statistical methods in epidemiology. These methods need to be tailored to the data available and problems they are attempting to address.

The SARS-CoV-2 pandemic highlighted the benefit of reliable predictive mathematical models and their use in determining the effect of non-pharmaceutical intervention

measures and management of medical resources. Outcomes from models can never be 100% certain, with assumptions and random events influencing model outcomes and true events, respectively. Having multiple models with different assumptions and structure, all attempting to tackle the same issue, can aid in consolidating results. It is important to highlight these assumptions and uncertainties when communicating results to stakeholders and the public. Much of the work during the SARS-CoV-2 pandemic involved tackling issues in real time. This highlights the advantage of working with mathematical models for "what if" situations. For example, if a vector-borne disease were to emerge in the UK the models and interventions used for the air-borne SARS-CoV-2 virus would be unsuitable, and again work on controlling the outbreak would be in real time. Having collected knowledge and designed models in order to have a plan in place before an outbreak occurs may allow stakeholders to get ahead of the epidemic curve and minimise the probability/magnitude of outbreaks.

Parameters such as the transmission rate are not static and depend on behaviour and climate, amongst other variables. At the beginning of an outbreak assumptions have to be made about these parameters, then as the outbreak progresses data can be gathered to fit these parameters for the given scenario. Similarly, the impact of interventions may be influenced by these variables. When estimating the effect on an intervention this uncertainty needs to be considered. Although mathematical models provide useful in-site to the epidemiological impact of interventions, policy makers also consider economic impact and politics in their decision making Klein et al. [2007]. The economic impact of interventions can be accessed from mathematical model results Balike Dieudonné [2021], Harris et al. [2016].

When considering spatial models different countries/communities have population aggregation, which will likely effect the transmission of some pathogens. This is also true for the distribution and transmission of livestock pathogens. Should countries be faced with the introduction of a foreign equine disease such as AHSV, the current uncertainty in spatial data is likely to effect the ability to accurately model the pathogens implications. Proactively gathering more informed data before an outbreak would allow the epidemiological and modelling tools available to be applied in in real-time.

Mathematical models can also be of use for understanding vector ecology [Ewing et al., 2019]. During the Moroccan AHSV outbreak there was no transmission across the Atlas Mountains. A possible explanation for this is the altitude acted as an effective barrier to *Culicoides* movements. Further studies of vector populations at different altitudes could help improve understanding of this phenomenon. Geographical regions in which vector species can survive are increasing with climate change [Wittmann and Baylis, 2000]. Also, if a region has not experienced an out break of a virus, often little in known of the potential of native species to transmit the virus. Laboratory studies of vector competence, such as those performed by Carpenter et al. [2011], can help inform mathematical models.

CHAPTER 2

Equine viral encephalitis: prevalence, impact, and management strategies

Note: This chapter is a published article:

Barba M, Fairbanks EL, Daly JM. Equine viral encephalitis: prevalence, impact, and management strategies. Veterinary Medicine: Research and Reports. 2019;10:99.

Author contribution: Formal analysis, Investigation, Visualization, Writing - original draft and Writing - review editing.

Veterinary Medicine: Research and Reports

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Equine viral encephalitis: prevalence, impact, and management strategies

This article was published in the following Dove Press journal:
Veterinary Medicine: Research and Reports

Marta Barba[1]
Emma L Fairbanks[2]
Janet M Daly[2]

[1]Veterinary Faculty, Universidad Cardenal Herrera-CEU, CEU Universities, Valencia, Spain; [2]School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, Leicestershire, UK

Abstract: Members of several different virus families cause equine viral encephalitis, the majority of which are arthropod-borne viruses (arboviruses) with zoonotic potential. The clinical signs caused are rarely pathognomonic; therefore, a clinical diagnosis is usually presumptive according to the geographical region. However, recent decades have seen expansion of the geographical range and emergence in new regions of numerous viral diseases. In this context, this review presents an overview of the prevalence and distribution of the main viral causes of equine encephalitis and discusses their impact and potential approaches to limit their spread.

Keywords: arbovirus, vector, vaccination, mathematical modeling

Introduction

Viral encephalitis is one of the most common infections of the central nervous system (CNS) in horses worldwide.[1] Clinical signs can include mild fever, dullness, sleepiness, listlessness, ataxia, inability to rise trembling, skin twitching, difficulty in urination and defecation, facial paralysis, blindness, seizures, coma, and other non-neurological signs.[1,2] The combination, severity, and duration of these clinical signs can vary depending on the etiological agent and its virulence; infection can be fatal. As clinical signs are usually very similar among the different diseases, which specific pathogen is considered depends on geographical areas. This review focuses on the main neurotropic viruses that cause encephalitis in equids and not viruses that can cause other neurological diseases such as equine herpes myeloencephalopathy. Other viruses that cause encephalitis in horses less frequently or affecting a smaller region are listed in Table 1.

Prevalence and distribution

Alphaviruses

Encephalitic alphaviruses belonging to the family *Togaviridae* cause neurological signs in equids and humans on the American continent.[3] The most common equine encephalitic viruses are Eastern equine encephalitis (EEE), Western equine encephalitis (WEE), and Venezuelan equine encephalitis (VEE). Collectively known as the equine encephalitides, they are transmitted by mosquitoes and wild birds are the main reservoir host. Horses and humans are considered dead-end hosts for EEE and WEE viruses because they do not generate enough viremia to infect mosquitoes and perpetuate the transmission cycle. On the other hand, equids are the key reservoir host for VEE virus because they develop high titer viremia that can act as source of infection for subsequent feeding mosquitoes.[2,4]

Correspondence: Janet M Daly
School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, Leicestershire LE12 5RD, UK
Tel +44 115 951 6475
Email janet.daly@nottingham.ac.uk

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/VMRR.S168227

Table 1 Other equine encephalitic viruses

Family	Genus	Virus	Geographical distribution	Reservoir host	Other hosts	Vector borne	Zoonotic potential	Reference
Togaviridae	Alphavirus	Highlands J	North America	Birds	Equids	Y	Y	84,85
		Ross River	Australia	Marsupials Horses Birds	Human	Y	Y	86,87
		Middleburg	Africa	Birds	Equids Ruminants Human	Y	Y	88–90
		Sindbis	Africa, Eurasia Australia	Birds	Equids Human	Y	Y	91
Flaviviridae	Flavivirus	Murray valley	Australia New Guinea	Birds	Equids Cattle Marsupials Fox	Y	Y	92
		Kunjin	Australia	Birds	Equids Human	Y	Y	92
		St. Louis encephalitis Usutu	North America Africa Europe	Birds Birds	Equids Human Human Ruminants Equids	Y Y	Y Y	93 90
		Louping ill	Spain Portugal UK	Sheep Grouse	Equids Human	Y	Y	94
		Powassan	North America Russia	Lagomorphs Rodents Skunks Dogs Birds	Equids Human	Y	Y	95
		Tick-borne encephalitis	Asia Europe Finland Russia	Small rodents	Equids Human Primates Dogs Ruminants	Y	Y	96–98

(*Continued*)

Table 1 (Continued).

Family	Genus	Virus	Geographical distribution	Reservoir host	Other hosts	Vector borne	Zoonotic potential	Reference
Bunyaviridae	Orthobunyaviridae	California Serogroup (California encephalitis, Jamestown Canyon, La Crosse, Snowshoe hare)	North America	Rodents Lagomorphs	Equids Human	Y	Y	99
		Shuni Virus	Africa	Ruminants	Equids Human	Y	Y	100
Reoviridae	Orbivirus	African horse sickness	Africa	Equids	-	Y	N	101
		Equine encephalosis	Africa	Equids Elephants	-	Y	N	102

EEE virus

In North America, EEE has been considered endemic for decades.[5] This disease is more prevalent in the Southeastern region of the United States with a high fatality rate. However, since 2005, the geographic range of the virus has spread northwards,[6,7] and 8.7% seroprevalence was reported in horses in southern Quebec in 2012.[8] Madariaga virus (MADV) is the new species designation for the South American isolates of EEE virus (previously referred to as EEE lineages II, III, and IV) to reflect the different pathogenesis and ecology and genetic divergence from North American strains.[9] In Central and South America, small outbreaks of MADV with low fatality rate have been reported between the1930s and 1990s.[5] More recently, larger outbreaks of higher morbidity and mortality have occurred.[10–12] In Brazil, high fatality rate outbreaks were reported between 2008 and 2009 with 229 horses affected.[13] In 2010, seroprevalence of MADV in horses was reported to be 26.3% in Panama and 9.9% in Brazil.[11,14]

WEE virus

In North America, the WEE virus has traditionally affected states west of the Mississippi river, with the largest outbreaks registered in the 1930s and 1940s in Canada and the United States affecting hundreds of thousands of equids.[2] However, no cases have been reported in North America since 1998 and the last time the virus was detected in mosquito pools was in 2008.[15]

In South and Central America, the last confirmed equine outbreak was reported in Brazil in 2007 and a prevalence of 36.4% has been reported in non-vaccinated horses in the Pantanal region of Brazil in 2010 and 0.4% in 2015.[14] The disease is suspected but has not been confirmed in other countries such as Bolivia and Costa Rica. In Uruguay, a fatal human case in 2009 associated with WEE virus encephalitis in a child led to a seroprevalence survey in this country, which revealed a low prevalence of this virus in horses ranging from 3% to 4%.[16,17]

WEE virus is an example of an apparently declining equine and human pathogen probably caused by a reduction in genetic diversity of circulating lineages, which contrasts with the recent emergence of other arboviruses.[15,18]

VEE virus

Generally, only the epizootic strains 1-AB and 1-AC of VEE virus produce encephalomyelitis in horses, with a fatality rate close to 90%.[19,20] Outbreaks of this disease started in South America and spread northward via Central

America up to Mexico and southern Texas affecting hundreds of thousands of horses.[21] In Mexico, a high equine seroprevalence has been reported ranging from 17% to 80% in different states between 2003 and 2010.[22]

Flaviviruses

The family *Flaviviridae* contains the largest number of viral species that may cause encephalitis in horses. All are zoonotic and transmitted by mosquitoes or ticks (Table 1). The most significant are West Nile virus (WNV) and Japanese encephalitis virus (JEV).[1]

West Nile virus

West Nile virus is the flavivirus with the widest distribution, which includes all continents (Table 2). In affected regions, WNV is maintained in an enzootic cycle between mosquitoes and birds.[23] Horses and humans are considered dead-end hosts because of the low viremia developed, which is not sufficient to transmit the virus back to mosquitoes.[3] Experimental studies have demonstrated that only 10% of the infected horses develop neurological signs, but they can be lethal.[24] WNV was first isolated in Africa in 1937 and spread to Eurasia and Australia where sporadic outbreaks occurred.[25] Since the 1990s, more frequent outbreaks have occurred in the Mediterranean Basin and WNV appeared for the first time in North America in 1999, subsequently spreading across the continent.[25,26] Since 2008, a re-emergence of WNV in Central and Southeastern Europe has been observed, with both lineage 1 and lineage 2 WNV involved in outbreaks.[27] In Australia, WNV was named Kunjin virus, which was endemic in northern Australia but has caused recent outbreaks of encephalitis in horses in the southeast probably because of enhanced vector transmission.[28,29]

Japanese encephalitis virus

Japanese encephalitis virus most commonly circulates amongst birds and mosquitoes.[30] Pigs are referred to as a virus-amplifying host because they develop high viremia.[31] As for EEE virus and WEE virus, horses and humans are dead-end hosts for JEV.[32] The virus is endemic in southern areas of Asia and some Pacific countries, such as Malaysia, Indonesia, Singapore, New Guinea and Australia where sporadic outbreaks are observed.[33] Whereas in northern Asiatic areas such as Korea, Nepal, China, Taiwan, Japan, northern parts of Vietnam, India or Thailand; seasonal epidemics develop.[33] In Korea, around half of 989 horses tested between 2005 and 2007 were antibody positive.[34] In India, 10.5% of 637 horses screened between 2006 and 2010 had antibodies against JEV.[35]

Mononegavirales

Viruses in the order Mononegavirales are large enveloped viruses with a single-stranded negative-sense RNA genome. Several families in the order (*Rhabdoviridae, Orthobornaviridae,* and *Paramyxoviridae*) include viruses that can produce encephalitis in animals and humans.[36]

Rabies virus

Rabies virus, a neurotropic virus in the genus *Lyssavirus* (family *Rhabdoviridae*), is one of the deadliest zoonoses worldwide.[37] European countries, Iceland, Greenland, New Zealand, and Australia are considered free of this disease, but it is present on the American, African, and Asian continents.[38] All mammals are susceptible, but canids and bats are the major vectors. Transmission is via saliva, mainly when a rabid animal bites another animal. Rabies infection is relatively rare in horses; only 23 rabid equids were reported in the United States in 2016 and 13 in 2017.[39] Nevertheless, in some African countries, large numbers of rabies cases occur in equids, including donkeys, and there may occasionally be transmission to people.[40]

Borna disease virus

Borna disease virus-1 (BoDV-1) is a neurotropic pathogen in the genus *Orthobornavirus* (family *Bornaviridae*) that causes mononuclear encephalomyelitis in horses.[41] This disease is endemic in certain areas in central Europe including Germany, Switzerland, Liechtenstein, and Austria and is usually fatal. The reservoir host of this virus is the bicolored white-toothed shrew (*Crocidura leucodon*), but natural infection can occur occasionally in equids and other animals such as sheep, cattle, llamas, cats, dogs, and ostriches.[42] A landscape modeling study conducted in an endemic area suggested that horses come into contact with shrews in dry habitats such as grasslands and stables.[43] In Germany, close to the town of Borna, large numbers of horses died in the late 1800s.[42,44] In the 1990s, the incidence was much lower, around 100 horses per year in the endemic area.[45] Recently, a new endemic area in Austria was reported after confirmation of lethal disease in horses.[46] One case has been reported in the United Kingdom in a horse imported from Germany.[47] Antibodies against Bornaviruses have been detected in equids in non-endemic areas of Europe, Iceland, Turkey, Israel, Japan, China, Iran, Australia, and United States.[48–53] However, it is not considered proof of infection due to the cross-reactivity with

Table 2 Recently published seroprevalence of West Nile virus in some countries

Country	Seroprevalence	Year	Test used	Reference
Algeria	17.4% (26.8% horses, 14.4% donkeys)	2014	ELISA confirmed with WB and VNT	103
Argentina	16.2%	2008	PRNT	104
Australia (KUNV)	4.8%	2011	cELISA confirmed by PRNT	105
Brazil	1.46%	2004–2009	ELISA and VNT	104
Canada	16.5%	2012–2014	ELISA confirmed by PRNT	106
Chad	97%	2003–2004		107
Cote d'Ivoire	28%	2003–2005		107
Croatia	3.43%	2010–2011	ELISA confirmed with VNT and PRNT	108
France	35%	2003	ELISA and VNT	109
Gabon	3%	2004		107
Greece	33%	2010	cELISA	110
Israel	84.6%	2014	cELISA and VNT	111
Italy	39.1%	2008	-	112
Mexico	26%	2006	cELISA	113
	45%	2007		
Morocco	31%	2011	ELISA and VNT	114
Pakistan	65%	2012–2013	cELISA (anti-pr-E IgG)	115
Palestine	48.6%	2014	cELISA	111
Poland	15.08%	2012–2013	VNT	116
Romania	58.5%	2010	cELISA	111
	15.2%	2006–2008		
Saudi Arabia	17.3–55.6% (depending on region)	2013–2015	ELISA and VNT	117
Senegal	92%	2002–2003	ELISA confirmed with PRNT	107
Slovak Republic	6.9%	2013	cELISA and NT	118
Spain	7.1% (CI 95% 5.4–11.2%)	2010	cELISA and VNT	119
Tunisia	28% (95% CI 22–34%)	2009	ELISA and VNT	120
Turkey	4.9–30.6% (depending on regions)	2011–2013	PRNT	121
Ukraine	13.5%	2010–2011	ELISA and PRNT	122
USA	19% (feral horses)[a]	2008	ELISA confirmed by PRNT	123
	7.2% (feral horses)	2009		
Venezuela	4.3%	2004–2006	ELISA confirmed with PRNT	124

Notes: [a]Widespread vaccination in horses in this country precludes performing seroprevalence studies.
Abbreviations: cELISA, competition ELISA; ELISA, enzyme-linked immunosorbent assay; PRNT, plaque reduction neutralization test; VNT, virus neutralization test.

avian Bornaviruses.[46] There are sporadic reports of confirmed human BoDV-1 infection including a recent fatal encephalitis case,[54] but it is unclear whether these represent interspecies transmission from horses or other hosts. An association between BoDV-1 infection and human neuropsychiatric disease was first reported in 1985,[55] although this remains controversial.

Hendra and Nipah virus

The name of in the genus Henipavirus (family *Paramyxoviridae*) is an amalgamation of Hendra and Nipah. Both species are emerging zoonotic pathogens for which flying foxes (bats in the genus Pteropus) are the reservoir host. Hendra virus (HeV) causes respiratory and often fatal neurological disease in horses and people. It emerged in Brisbane in 1994 and is restricted to Australia.[56] Prevalence is low as most cases occur as spillover events to individual horses. There is a risk of human transmission during the preclinical stages of the disease and all infected people had close direct contact with body fluids from infected horses.[57]

Nipah virus (NiV), which has circulated in Malaysia and Singapore since the late 1990s, has spread to Thailand, India, and Bangladesh.[58] It mainly affects domestic pigs

and people but can occasionally affect horses producing encephalitis and meningitis.[59]

Prevention and control

The majority of equine encephalitic viruses are limited to specific geographical areas. Spread to disease-free areas of the world can have catastrophic consequences on equine welfare and industry including mortality, loss of earnings, increased costs (due to veterinary treatment and hospitalization, and preventive measures such as vaccination), as well as public health consequences. For example, outbreaks of African horse sickness in the past have caused 300,000 equine deaths in a short time. It was estimated that the economic cost of such an outbreak in the Netherlands could be more than 500 million Euros.[60] In another recent study, it was estimated that the cost of a WNV epidemic in Belgium would be over 30 million euros for equine patients and over 45 million euros for human patients.[61]

Viral outbreaks are not completely avoidable, but preventative strategies can help restrict their occurrence. Management strategies can also be used in an attempt to eradicate a pathogen from a population or limit its impact. Particularly due to its zoonotic potential, an outbreak of HeV led to the re-evaluation of infection control and equine management practices in Queensland, Australia. Horses can also be used as epidemiological sentinels for human surveillance.[2] For example, although horses are not believed to be an amplifying host of EEE virus epidemics, they tend to be the first to show clinical signs, therefore providing the first indication that the virus is circulating.[4] Thus, illness detection in horses can trigger measures to prevent associated outbreaks in humans. Viruses can be spread through many different mechanisms, therefore warranting different control strategies.[62] New equine encephalitic viruses are still being discovered, for example, HeV and NiV were both first identified in the 1990s,[63] and there are likely to be more that remain undiscovered. Control of future emerging virus outbreaks may rely on identifying appropriate strategies already applied to related known diseases. Mathematical modeling can provide an understanding of mechanisms driving disease outbreaks. However, it is important to consider how reliable the values assigned to parameters are ("parameter identifiability") before mathematical models are used to guide interventions.

Diagnostic techniques

The increasing threat of vector-borne diseases emphasizes the importance of vector surveillance systems and diagnostic tests for early detection of pathogens.[64] Early identification of the virus causing equine encephalitis will improve the effectiveness of many disease control measures. As previously mentioned, clinical diagnosis of equine encephalitic viruses is often unreliable due to overlap in the clinical signs seen; therefore, laboratory testing is usually necessary to confirm the etiology of the disease. The OIE (World Organisation for Animal Health) Manual of Diagnostic Tests and Vaccines for Terrestrial Animals describes internationally agreed diagnostic tests for each of the virus species presented in this review with the exception of BoDV-1. The preferred diagnostic test varies for the different viruses and the purpose for which it is being performed, which can include confirmation of a clinical case, surveillance, demonstrating freedom from infection of an individual animal or population and monitoring the response to vaccination. Virus isolation can be time-consuming and for many of the viruses described requires high levels of laboratory containment, but the OIE recommends it as a definitive diagnostic of VEE virus. Polymerase chain reaction (PCR)-based techniques are widely used for virus detection as they offer the advantages of being specific and rapid to perform. However, for some viruses, particularly the flaviviruses, the transient nature of viremia means that RT-PCR tests frequently return false-negative results. Therefore, serological confirmation is necessary. Enzyme-linked immune-sorbent assays (ELISA) are increasingly popular as a relatively inexpensive and rapid diagnostic test. However, cross-reactivity between closely related co-circulating viruses complicates serological testing, particularly for flaviviruses.[6] Therefore, confirmatory testing using a virus-neutralizing test such as the plaque-reduction neutralization test is often required. Disease surveillance often includes random testing of animals in order to observe whether a pathogen is present within a population,[62] which also requires cost-effective and accurate assays to be developed.

Vaccination

Vaccines are currently available for many of the viruses that cause equine encephalitis (Table 3). An equine vaccine for JEV is notably missing although human vaccines are available and are sometimes administered to horses (eg, in Japan). In contrast, although there are equine vaccines against WNV, there is no human WNV vaccine. Although live-attenuated and inactivated virus vaccines have successfully prevented disease for many decades, these vaccines have some limitations. For example, inactivated virus vaccines typically induce short-lived protective antibody responses and there is a risk of

Table 3 Vaccines licensed for use in horses to protect against viruses that cause encephalitis

Virus	Vaccine type
Eastern equine encephalitis	Inactivated whole virus
Western equine encephalitis	Inactivated whole virus
Venezuelan equine encephalitis	Inactivated whole virus
(VEE)	A conditionally available modified live virus (MLV) VEE vaccine has been released during previous outbreaks
West Nile	Inactivated whole virus
	Modified live (canarypox vector expressing prM and E proteins)
	DNA vaccine
Rabies	Inactivated whole virus
Hendra	Subunit (recombinant glycoprotein)

Abbreviations: prM, membrane, E, envelope.

reversion to virulence with inactivated virus vaccines. This has led to the development of second-generation vaccines, such as the live-vectored and DNA vaccines available for WNV. These vaccines often enable a "differentiation of infected from vaccinated animals" (DIVA) approach to be taken whereby diagnostic tests are used that detect antibodies against proteins not generated in response to the vaccine. This can be critical in controlling an emerging virus outbreak as it enables authorities to determine when an outbreak is over by screening for antibodies that only develop in infected animals.

Vaccination coverage is often determined by factors such as economic and logistic issues in developing countries and motivational and legislative issues in developed countries.[65] Mass vaccination is not likely to be cost-effective; focusing on high-risk groups would most likely be more appropriate.[64] Furthermore, it is not always necessary to vaccinate every individual for a population to be protected. The basic reproduction number, R_0, is the expected number of new infectious cases generated from an individual host during their infectious period. When this value is larger than one (R_0>1), we expect the number of infected individuals to increase, and if R_0 is less than one, we expect the disease to die out of the population. Considering this, it is possible to approximate the proportion of a population that require vaccination in order to stop the pathogen circulating, therefore reducing R_0 to below unity. By vaccinating a proportion (p) of the population, the R_0 is decreased to $(1-p)R_0$. This allows derivation of a condition for this proportion; as $(1-p)R_0$ must be less than one, we have

$$p>1-\frac{1}{R_0}$$

This shows that it is not necessary to vaccinate the whole population, as unvaccinated horses will be protected from the vaccination of others, known as herd immunity.[65] Empirical studies have confirmed this theoretical idea.[66] Vaccination has led to the global eradication of smallpox and rinderpest virus. However, herd immunity and disease eradication are more difficult to achieve for viruses with reservoir host species or insect vectors.

Control of exposure to viral vectors and reservoir hosts

Reducing exposure of horses to wildlife that transmit equine encephalitic viruses can be difficult to achieve. Population control methods such as vaccination and/or sterilization of wild or feral canids have been widely employed to reduce human transmission of rabies.[67] However, this approach can cause ethical debate, for example, where poisoning of bats has been used to control rabies in South America.

At the equine premises level, exposure to insect vectors can be reduced by using fly rugs and insect repellents, and stabling horses during peak vector activity (eg, at dusk).[62] Other localized methods of vector control include mass trapping and blocking breeding sites by obstructing water surfaces with polystyrene balls.[68] Control measures also include reducing mosquito populations.[25] The use of pesticides to control vector-borne viruses raises environmental and health concerns and mosquito populations are developing resistance to conventional control agents. There has been an increased interest in the development of biopesticides[68] and the creation of genetically modified mosquitoes that cannot transmit pathogens.[69–71] Mathematical modeling has predicted that if the abundance of mosquitoes could be reduced such that R_0 becomes <1then WNV would die out.[72] Wonham et al (2004) predicted that if the initial size of the New York mosquito population was 40–70% smaller, the outbreak of WNV in 2000 could have been prevented.[73] In contrast, reducing the bird population increases the chance of an

outbreak as it increases the ratio of mosquitoes to birds making virus transmission more likely, as long as the population is not reduced to the extent that the mosquito population is not maintained. Although mosquitoes are the most common insect vector of equine encephalitic viruses, ticks (eg, Powassan virus[74] and louping ill virus) and midges (eg, African horse sickness viruses) can also act as vectors.[3]

Control of disease spread through international movement of horses

Increased globalization has led to a greater potential for the spread of infectious diseases. Most international equine movements are for competition purposes. The number of prestigious international competition events has increased in the last 10–15 years. There has also been an increase in the number of stallions being transported between the northern and southern hemispheres for breeding; this number rose from 7 in 1989 to over 100 in 2000.[63] In addition, horses may be transported as a result of change of ownership or slaughter in the meat industry. Countries often have different strategies of restricting pathogens from entering, including testing and quarantine of imported animals. They may also place restrictions on importation from specific countries to prevent introduction of certain pathogens. However, this can have an impact on the equine industry given the frequent international movement of some horse populations.[62] Quarantine, disinfection, pathogen screening, and transport restrictions are useful tools in infection control and biosecurity systems; however, these require optimization for maximum impact.[62]

It is not possible to predict when a new virus (either a newly identified pathogen or a known pathogen in a new geographical area) will emerge. However, models to assess the risk of a virus entering a population and changes in risk over time can be developed. There are also models for the risk of disease introduction through host movement.[76,77] Countries can be characterized as high risk (virus circulating), low risk (previous outbreaks and/or main vector present), and very low risk. If different host species or reasons for travel are associated with different risk levels then these can be further subdivided into groups. Risk pathways can then be constructed for the steps required for incursion (Figure 1). From these pathways, stochastic risk models that quantify the risk that importation of different groups of animals can be developed. These methods allow us to assess control strategies such as quarantine and their effectiveness on different risk groups.

In the case of vector-borne viruses, seasonal prevalence and vector abundance in endemic regions and regions at risk of disease introduction can be taken into account.

Vector-borne diseases are often restricted to temperate climates due to the range of the insects.[62] However, with climate change, the areas inhabited by virus-transmitting insects are changing.[78,79] The spread of vector-borne viruses is strongly influenced by temperature. Temperature has an effect on the life cycle of the insects, as well as the extrinsic incubation period (the time between a vector acquiring an infectious agent and becoming infectious). The average global temperature is predicted to increase between 1°C and 4.6°C during this century.[75] Increased temperatures and altered rainfall patterns are likely to affect the range and behavior of insect vectors.[63] Access to breeding sites also has an effect on the distribution of mosquitoes; an increased transmission of EEE virus has been associated with the freshwater hardwood swamps in the Atlantic and Gulf Coast states and the Great Lakes region (USA).[4] This is important to consider in the case of vector-borne diseases, as even if an infected host enters a naïve population, the virus cannot spread without the presence of its vector. The main species of vector, geographical distribution, and zoonotic potential vary between equine encephalitic viruses.[4] Geographic Information System-based spatial models for predicting locations with high risk have been developed;[68,80–82] these use predictor variables such as temperature, rainfall, and landscape/vegetation.[25]

Whereas some regions may be able to support the vector life cycle throughout the year, viruses may overwinter in unidentified hosts or be re-introduced (eg, by importation or migratory birds) in some climates. Some viruses, such as WNV, are transmitted vertically (from adults to eggs) within mosquito populations; this provides a mechanism for the viruses to be maintained within the population.[64] Vector-

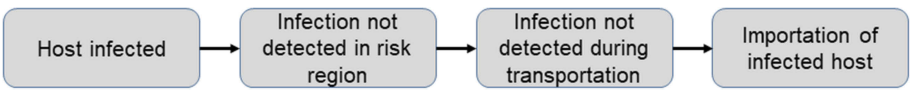

Figure 1 Pathway of the steps required for the incursion of a virus due to importation of infected host.

borne diseases can spread when vectors are carried by the wind. Incursions of JEV into northeastern Queensland are thought to be most likely due to infected mosquitoes blown by the wind from Papua New Guinea.[63]

International horse movements are not only a threat to naïve populations into which a new pathogen is introduced; the imported equine is also at risk of acquiring disease. An example of this occurred in horses imported to Korea from Ireland, New Zealand, and Australia in 1996 that became infected with JEV.[63,83] This highlights the importance of vaccinating horses against viruses they may come into contact before they travel, for example, horses that travel from the United Kingdom may be vaccinated against WNV.[63]

Conclusion

There is an apparent general increase in viral emergence and re-emergence, particularly of arboviruses. This trend includes viruses that cause potentially devastating encephalitic disease in horses. As a result, there is increasing awareness of the need to monitor disease trends in equine populations, particularly of viruses with zoonotic potential, and to formulate approaches to prevent or control disease outbreaks.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Long MT. West nile virus and equine encephalitis viruses. *Vet Clin North Am Equine Pract*. 2014;30:523–542. doi:10.1016/j.cveq.2014.08.001
2. Kumar B, Manuja A, Gulati B, Virmani N, Tripathi BN. Zoonotic viral diseases of equines and their impact on human and animal health. *TOVJ*. 2018;12:80–98. doi:10.2174/1874357901812010080
3. Chapman GE, Baylis M, Archer D, Daly JM. The challenges posed by equine arboviruses. *Equine Vet J*. 2018;50:436–445. doi:10.1111/evj.12829
4. Zacks MA, Paessler S. Encephalitic alphaviruses. *Vet Microbiol*. 2010;140(3–4):281–286. doi:10.1016/j.vetmic.2009.08.023
5. Brault AC, Powers AM, Villarreal Chavez CL, et al. Genetic and antigenic diversity among eastern equine encephalitis viruses from North, Central, and South America. *Am J Trop Med Hyg*. 1999;61:579–586. doi:10.4269/ajtmh.1999.61.579
6. Graham AC, Ledermann JP, Saxton-Shaw KD, et al. The first outbreak of Eastern equine encephalitis in Vermont: outbreak description and phylogenetic relationships of the virus isolate. *PLoS One*. 2015;10:e0128712. doi:10.1371/journal.pone.0128712
7. Rocheleau JP, Arsenault J, Ogden NH, Lindsay LR, Drebot M, Michel P. Characterizing areas of potential human exposure to eastern equine encephalitis virus using serological and clinical data from horses. *Epidemiol Infect*. 2017;145:667–677. doi:10.1017/S0950268817000668
8. Rocheleau J-P, Arsenault J, Lindsay LR, et al. Eastern equine encephalitis virus: high seroprevalence in horses from Southern Quebec, Canada, 2012. *Vector Borne Zoonotic Dis*. 2013;13:712–718. doi:10.1089/vbz.2012.0964
9. Riet-Correa F, Nogueira ML, VMK M, et al. Isolation and characterization of madariaga virus from a Horse in Paraíba State, Brazil. *Transbound Emerg Dis*. 2015;64:990–993. doi:10.1111/tbed.12382
10. Aguilar PV, Robich RM, Turell MJ, et al. Endemic eastern equine encephalitis in the Amazon region of Peru. *Am J Trop Med Hyg*. 2007;76:293–298. doi:10.4269/ajtmh.2007.76.293
11. Carrera J-P, Forrester N, Wang E, et al. Eastern equine encephalitis in Latin America. *New Engl J Med*. 2013;369:732–744. doi:10.1056/NEJMoa1212628
12. de Novaes Oliveira R, Iamamoto K, Silva MLCR, et al. Eastern equine encephalitis cases among horses in Brazil between 2005 and 2009. *Arch Virol*. 2014;159:2615–2620. doi:10.1007/s00705-014-2121-4
13. Silva MLCR, Galiza GJN, Dantas AFM, et al. Outbreaks of eastern equine encephalitis in northeastern Brazil. *J Vet Diagn Invest*. 2011;23:570–575. doi:10.1177/1040638711403414
14. Pauvolid-Corrêa A, Soares Juliano R, Campos Z, Velez J, Nogueira RMR, Komar N. Neutralising antibodies for mayaro virus in Pantanal, Brazil. *Mem Inst Oswaldo Cruz*. 2015;110:125–133. doi:10.1590/0074-02760140383
15. Negi SS, Braun WA, Auguste AJ, Bergren NA, Forrester NL, Weaver SC. Western equine encephalitis virus: evolutionary analysis of a declining alphavirus based on complete genome sequences. *J Virol*. 2014;88:9260–9267. doi:10.1128/JVI.01463-14
16. Burgueño A, Frabasile S, Díaz LA, et al. Genomic characterization and seroprevalence studies on alphaviruses in Uruguay. *Am J Trop Med Hyg*. 2018;98:1811–1818. doi:10.4269/ajtmh.17-0980
17. Delfaro A, Burgueño A, Morel N, et al. Fatal human case of Western Equine Encephalitis, Uruguay. *Emerg Infect Dis*. 2011;17:952–954. doi:10.3201/eid1701.100876
18. Kading RC, Hartman DA, Bergren NA, et al. Continued evidence of decline in the enzootic activity of western equine encephalitis virus in Colorado. *J Med Entomol*. 2018;1987:1–5.
19. Estrada-Franco JG, Carrara A-S, Aronson JF, Weaver SC, Gonzalez D. Equine amplification and virulence of subtype IE venezuelan equine encephalitis viruses isolated during the 1993 and 1996 Mexican Epizootics. *Emerg Infect Dis*. 2012;9:162–168.
20. Forrester NL, Wertheim JO, Dugan VG, et al. Evolution and spread of Venezuelan equine encephalitis complex alphavirus in the Americas. *PLoS Negl Trop Dis*. 2017;11:1–19. doi:10.1371/journal.pntd.0005693
21. Navarro-Lopez R, Clements T, Freier JE, et al. Venezuelan equine encephalitis virus, Southern Mexico. *Emerg Infect Dis*. 2012;10:2113–2121.
22. Adams AP, Navarro-Lopez R, Ramirez-Aguilar FJ, et al. Venezuelan equine encephalitis virus activity in the Gulf Coast region of Mexico, 2003-2010. *PLoS Negl Trop Dis*. 2012;6:2003–2010. doi:10.1371/journal.pntd.0001875
23. David S, Abraham AM. Epidemiological and clinical aspects on West Nile virus, a globally emerging pathogen. *Infect Dis (Lond)*. 2016;48:571–586. doi:10.3109/23744235.2016.1164890
24. Bunning ML, Bowen RA, Cropp CB, et al. Experimental infection of horses with West Nile virus. *Emerg Infect Dis*. 2002;8:380–386. doi:10.3201/eid0804.010239
25. Kramer LD, Styer LM, Ebel GD. A global perspective on the epidemiology of West Nile virus. *Annu Rev Entomol*. 2008;53:61–81.
26. Hofmeister EK. West Nile virus: north American experience. *Integr Zool*. 2011;6:279–289. doi:10.1111/j.1749-4877.2011.00251.x
27. Bakonyi T, Ivanics E, Erdelyi K, et al. Lineage 1 and 2 strains of encephalitic West Nile virus, central Europe. *Emerg Infect Dis*. 2006;12:618–623. doi:10.3201/eid1204.051379
28. van den Hurk AF, Hall-Mendelin S, Webb CE, et al. Role of enhanced vector transmission of a new West Nile virus strain in an outbreak of equine disease in Australia in 2011. *Parasit Vectors*. 2014;7:586. doi:10.1186/1756-3305-7-23

29. Prow NA, Edmonds JH, Williams DT, et al. Virulence and evolution of West Nile Virus, Australia, 1960-2012. *Emerg Infect Dis*. 2016;22:1353–1362. doi:10.3201/eid2208.151719
30. Kumar K, Arshad SS, Selvarajah GT, et al. Prevalence and risk factors of Japanese encephalitis virus (JEV) in livestock and companion animal in high-risk areas in Malaysia. *Trop Anim Health Prod*. 2018;50:741–752. doi:10.1007/s11250-017-1490-6
31. Ricklin ME, Garcia-Nicolas O, Brechbuhl D, et al. Japanese encephalitis virus tropism in experimentally infected pigs. *Vet Res*. 2016;47:34. doi:10.1186/s13567-016-0319-z
32. Mansfield KL, Hernandez-Triana LM, Banyard AC, Fooks AR, Johnson N. Japanese encephalitis virus infection, diagnosis and control in domestic animals. *Vet Microbiol*. 2017;201:85–92. doi:10.1016/j.vetmic.2017.01.014
33. Wang H, Liang G. Epidemiology of Japanese encephalitis: past, present, and future prospects. *Ther Clin Risk Manag*. 2015;11:435–448.
34. Yang DK, Kim BH, Kweon CH, et al. Serosurveillance for Japanese encephalitis, Akabane, and Aino viruses for Thoroughbred horses in Korea. *J Vet Sci*. 2008;9:381–385. doi:10.4142/jvs.2008.9.4.381
35. Gulati BR, Singha H, Singh BK, Virmani N, Kumar S, Singh RK. Isolation and genetic characterization of Japanese encephalitis virus from equines in India. *J Vet Sci*. 2012;13:111–118. doi:10.4142/jvs.2012.13.2.111
36. Ebihara H, Tomonaga K, Calisher CH, et al. Taxonomy of the order Mononegavirales: update 2017. *Arch Virol*. 2017;162:2493–2504. doi:10.1007/s00705-017-3383-4
37. Barecha CB, Girzaw F, Kandi RV, Pal M. Epidemiology and Public Health Significance of Rabies. *Perspect Clin Res*. 2017;5:55–67.
38. Rupprecht CE, Willoughby R, Slate D. Current and future trends in the prevention, treatment and control of rabies. *Expert Rev Anti Infect Ther*. 2006;4:1021–1038. doi:10.1586/14787210.4.6.1021
39. Ma X, Monroe B, Cleaton J, Orciary L, Li Y, Al. E. Rabies surveillance in the United States during 2017. *J Am Vet Med Assoc*. 2018;253:1555–1568. doi:10.2460/javma.253.8.980
40. Jemberu WT, Molla W, Almaw G, Alemu S. Incidence of rabies in humans and domestic animals and people's awareness in North Gondar Zone, Ethiopia. *PLoS Negl Trop Dis*. 2013;7:e2216–e2216. doi:10.1371/journal.pntd.0002216
41. Tizard I, Ball J, Stoica G, Payne S. The pathogenesis of bornaviral diseases in mammals. *Anim Health Res Rev*. 2016;17:92–109. doi:10.1017/S1466252316000062
42. Lipkin WI, Briese T, Hornig M. Borna disease virus - Fact and fantasy. *Virus Res*. 2011;162:162–172. doi:10.1016/j.virusres.2011.09.036
43. Hermes N, Eickmann M, Encarnação JA, Herzog S, Becker NI, Herden C. Landscape features and reservoir occurrence affecting the risk for equine infection with Borna disease virus. *J Wildl Dis*. 2013;49:860–868. doi:10.7589/2012-10-262
44. Ludwig H, Bode L. Borna disease virus: new aspects on infection, disease, diagnosis and epidemiology Introduction and historical development. *Rev Sci Tech*. 2000;19:259–288.
45. Staeheli P, Sauder C, Hausmann J, Schwemmle M, Ehrensperger F. Epidemiology of Borna disease virus. *J Gen Virol*. 2000;81:2123–2135. doi:10.1099/0022-1317-81-9-2123
46. Weissenböck H, Bagó Z, Kolodziejek J, et al. Infections of horses and shrews with Bornaviruses in Upper Austria: A novel endemic area of Borna disease. *Emerg Microbes Infect*. 2017;6:1–9. doi:10.1038/emi.2017.36
47. Priestnall SL, Schoniger S, Ivens PAS, et al. Borna disease virus infection of a horse in Great Britain. *Vet Rec*. 2011;168:380. doi:10.1136/vr.c6405
48. Björnsdóttir S, Agustsdóttir E-L, Blomström A-L, et al. Serological markers of Bornavirus infection found in horses in Iceland. *Acta Vet Scand*. 2013;55:77. doi:10.1186/1751-0147-55-77
49. Inoue Y, Yamaguchi K, Sawada T, Rivero JC, Horii Y. Demonstration of continuously seropositive population against Borna disease virus in Misaki Feral horses, a japanese strain: a four-year follow-up study from 1998 to 2001. *J Vet Med Sci*. 2002;64:445–448. doi:10.1292/jvms.64.445
50. Kinnunen PM, Billich C, Ek-Kommonen C, et al. Serological evidence for Borna disease virus infection in humans, wild rodents and other vertebrates in Finland. *J Clin Virol*. 2007;38:64–69. doi:10.1016/j.jcv.2006.10.003
51. Pisoni G, Nativi D, Bronzo V, Codazza D. Sero-epidemiological study of Borna disease virus infection in the Italian equine population. *Vet Res Comm*. 2007;31:245–248. doi:10.1007/s11259-007-0016-5
52. Zhang L, Wang X, Zhan Q, et al. Evidence for natural Borna disease virus infection in healthy domestic animals in three areas of western China. *Arch Virol*. 2014;159:1941–1949. doi:10.1007/s00705-013-1971-5
53. Dauphin G, Legay V, Pitel P, Zientara S. Borna disease: current knowledge and virus detection in France. *Vet Res*. 2002;33:127–138. doi:10.1051/vetres:2002002
54. Korn K, Coras R, Bobinger T, et al. Fatal encephalitis associated with Borna disease virus 1. *New Engl J Med*. 2018;379:1375–1377. doi:10.1056/NEJMc1800724
55. Rott R, Herzog S, Fleischer B, et al. Detection of serum antibodies to Borna disease virus in patients with psychiatric disorders. *Science*. 1985;228:755–756. doi:10.1126/science.3922055
56. Middleton D. Hendra virus. *Vet Clin North Am Eq Pract*. 2014;30:579–589. doi:10.1016/j.cveq.2014.08.004
57. Marsh G, Haining J, Hancock TJ, et al. Experimental infection of horses with Hendra virus/Australia/horse/2008/Redlands. *Emerg Infect Dis*. 2011;17:2232–2238. doi:10.3201/eid1712.111162
58. Daszak P, Plowright RK, Epstein JH, et al. The emergence of Nipah and Hendra virus: pathogen dynamics across a wildlife-livestock-human continuum. In: Collinge SK, Ray C, editors. *Disease Ecology: Community Structure and Pathogen Dynamics*. Oxford: Oxford University Press; 2006:186–201.
59. Hooper P, Zaki S, Daniels P, Middleton D. Comparative pathology of the diseases caused by Hendra and Nipah viruses. *Microbes Infect*. 2001;3:315–322.
60. Robin M, Page P, Archer D, Baylis M. African horse sickness: the potential for an outbreak in disease-free regions and current disease control and elimination techniques. *Equine Vet J*. 2016;48:659–669. doi:10.1111/evj.12600
61. Humblet M, Vandeputte S, Fecher-Bourgeois F, et al. Estimating the economic impact of a possible equine and human epidemic of West Nile virus infection in Belgium. *Euro Surveill*. 2016;21(31):pii=30309. doi:10.2807/1560-7917.ES.2016.21.31.30309
62. Weese JS. Infection control and biosecurity in equine disease control. *Equine Vet J*. 2014;46:654–660. doi:10.1111/evj.12295
63. Timoney PJ. Factors influencing the international spread of equine diseases. *Vet Clin North Am Equine Pract*. 2000;16:537–551.
64. Kaaijk P, Luytjes W. Are we prepared for emerging flaviviruses in Europe? Challenges for vaccination. *Hum Vaccin Immunother*. 2018;14:337–344. doi:10.1080/21645515.2017.1389363
65. Nokes DJ, Anderson RM. The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. *Epidemiol Infect*. 1988;101:1–20. doi:10.1017/s0950268800029186
66. Anderson RM, May RM. The invasion, persistence and spread of infectious diseases within animal and plant communities. *Philos Trans R Soc Lond B Biol Sci*. 1986;314:533–570. doi:10.1098/rstb.1986.0072
67. Zinsstag J, Lechenne M, Laager M, et al. Vaccination of dogs in an African city interrupts rabies transmission and reduces human exposure. *Sci Transl Med*. 2017;9:eaaf6984. doi:10.1126/scitranslmed.aaf6984

68. Curtis C. Insecticide-treated nets against malaria vectors and polystyrene beads against Culex larvae. *Trends Parasitol*. 2005;21:504–507. doi:10.1016/j.pt.2005.08.025
69. Coutinho-Abreu IV, Zhu KY, Ramalho-Ortigao M. Transgenesis and paratransgenesis to control insect-borne diseases: current status and future challenges. *Parasitol Int*. 2010;59:1–8. doi:10.1016/j.parint.2009.10.002
70. Lacey LA, Frutos R, Kaya HK, Vail P. Insect pathogens as biological control agents: do they have a future? *Biol Control*. 2001;21:230–248. doi:10.1006/bcon.2001.0938
71. Sinkins SP, Gould F. Gene drive systems for insect disease vectors. *Nature Rev Genet*. 2006;7:427–435. doi:10.1038/nrg1870
72. Bowman C, Gumel AB, van den Driessche P, Wu J, Zhu H. A mathematical model for assessing control strategies against West Nile virus. *Bull Math Biol*. 2005;67:1107–1133. doi:10.1016/j.bulm.2005.01.002
73. Wonham MJ, de-Camino-Beck T, Lewis MA. An epidemiological model for West Nile virus: invasion analysis and control applications. *Proc Biol Sci*. 2004;271:501–507. doi:10.1098/rspb.2003.2608
74. Keane DP, Little PB. Equine viral encephalomyelitis in Canada: a review of known and potential causes. *Can Vet J*. 1987;28:497–504.
75. Ecology WB. How climate change alters rhythms of the wild. *Science*. 2000;287:793–795. doi:10.1126/science.287.5454.793
76. de Vos CJ, Hoek CA, Nodelijk G. Risk of introducing African horse sickness virus into the Netherlands by international equine movements. *Prev Vet Med*. 2012;106(2):108–122. doi:10.1016/j.prevetmed.2012.01.019
77. Faverjon C, Leblond A, Hendrikx P, et al. A spatiotemporal model to assess the introduction risk of African horse sickness by import of animals and vectors in France. *BMC Vet Res*. 2015;11:127. doi:10.1186/s12917-015-0435-4
78. Githeko AK, Lindsay SW, Confalonieri UE, Patz JA. Climate change and vector-borne diseases: a regional analysis. *Bull WHO*. 2000;78:1136–1147.
79. Reiter P. Climate change and mosquito-borne disease: knowing the horse before hitching the cart. *Rev Sci Tech*. 2008;27:383–398.
80. Kelen PV, Downs JA, Unnasch T, Stark L. A risk index model for predicting eastern equine encephalitis virus transmission to horses in Florida. *Appl Geog*. 2014;48:79–86. doi:10.1016/j.apgeog.2014.01.012
81. Moncayo AC, Edman JD, Finn JT. Application of geographic information technology in determining risk of eastern equine encephalomyelitis virus transmission. *AMCA*. 2000;16:28–35.
82. Zou L, Miller SN, Schmidtmann ET. A GIS tool to estimate West Nile virus risk based on a degree-day model. *Environ Monit Assess*. 2007;129:413–420. doi:10.1007/s10661-006-9373-8
83. University of Kentucky. *Equine Disease*. 1997. Available from: https://gluck.ca.uky.edu/sites/gluck.ca.uky.edu/files/q_oct_1997.pdf. Accessed July 8, 2019.
84. Borland EM, Ledermann JP, Powers AM. Culex Tarsalis Mosquitoes as Vectors of Highlands J Virus. *Vector-Borne Zoonotic Dis*. 2016;16:558–565. doi:10.1089/vbz.2015.1907
85. Karabatsos N, Lewis AL, Calisher CH, Hunt AR, Roehrig JT. Identification of Highlands J virus from a Florida horse. *Am J Trop Med Hyg*. 1988;39:603–606. doi:10.4269/ajtmh.1988.39.603
86. Gummow B, Tan RHH, Joice RK, Burgess G, Picard J. Seroprevalence and associated risk factors of mosquito-borne alphaviruses in horses in northern Queensland. *Aust Vet J*. 2018;96:243–251. doi:10.1111/avj.12711
87. Stephenson EB, Peel AJ, Reid SA, Jansen CC, McCallum H. The non-human reservoirs of Ross River virus: A systematic review of the evidence. *Parasit Vectors*. 2018;11:1–13. doi:10.1186/s13071-017-2573-y
88. Attoui H, Sailleau C, Jaafar FM, et al. Complete nucleotide sequence of Middelburg virus, isolated from the spleen of a horse with severe clinical disease in Zimbabwe. *J Gen Virol*. 2007;88:3078–3088. doi:10.1099/vir.0.83076-0
89. Burt FJ, Goedhals D, Mathengtheng L. Arboviruses in southern Africa: are we missing something? *Future Virol*. 2014;9:993–1008. doi:10.2217/fvl.14.87
90. Venter M. Assessing the zoonotic potential of arboviruses of African origin. *Curr Opin Virol*. 2018;28:74–84. doi:10.1016/j.coviro.2017.11.004
91. Niekerk SV, Human S, Williams J, Wilpe EV, Pretorius M, Swanepoel R. Neurologic disease in horses, South Africa. *Emerg Infect Dis*. 2015;21:2225–2229. doi:10.3201/eid2112.150132
92. Boyle DB, Dickerman RW, Marshall ID. Primary viraemia responses of herons to experimental infection with Murray Valley encephalitis, Kunjin and Japanese encephalitis viruses. *Austr J Exp Biol Med Sci*. 1983;61:655–664. doi:10.1038/icb.1983.62
93. White GS, Symmes K, Sun P, et al. Reemergence of St. Louis encephalitis virus, California, 2015. *Emerg Infect Dis*. 2016;22:2185–2188. doi:10.3201/eid2212.160805
94. Hyde J, Nettleton P, Marriott L, Willoughby K. Louping ill in horses. *Vet Rec*. 2007;160:532. doi:10.1136/vr.160.4.105
95. Little PB, Thorsen J, Moore W, Weninger N. Powassan viral encephalitis: a review and experimental studies in the horse and rabbit. *Vet Pathol*. 1985;22:500–507.
96. Rieille N, Klaus C, Hoffmann D, Péter O, Voordouw MJ. Goats as sentinel hosts for the detection of tick-borne encephalitis risk areas in the Canton of Valais, Switzerland. *BMC Vet Res*. 2017;13:1–13. doi:10.1186/s12917-016-0931-1
97. Rushton JO, Lecollinet S, Hubálek Z, Svobodová P, Lussy H, Nowotny N. Tick-borne encephalitis virus in horses, Austria, 2011. *Emerg Infect Dis*. 2013;19:635–637. doi:10.3201/eid1904.121450
98. Yoshii K, Song JY, Park SB, Yang J, Schmitt HJ. Tick-borne encephalitis in Japan, Republic of Korea and China. *Emerg Microbes Infect*. 2017;6:e82.
99. Beaty BJ, Bishop DHL. Bunyavirus-vector interactions. *Virus Res*. 1988;10:289–301.
100. van Eeden C, Williams JH, Gerdes TGH, et al. Shuni virus as cause of neurologic disease in horses. *Emerg Infect Dis*. 2012;18:318–321. doi:10.3201/eid1802.111403
101. Carpenter S, Mellor PS, Fall AG, Garros C, Venter GJ. African horse sickness virus: history, transmission, and current status. *Annu Rev Entomol*. 2017;62:343–358. doi:10.1146/annurev-ento-031616-035010
102. Dhama K, Pawaiya R, Karthik K, Chakraborty S, Tiwari R, Verna A. Equine encephalosis virus (EEV): a review. *AJAVA*. 2014;9:123–133.
103. Lafri I, Prat CM, Bitam I, et al. Seroprevalence of West Nile virus antibodies in equids in the North-East of Algeria and detection of virus circulation in 2014. *Comp Immunol Microbiol Infect Dis*. 2017;50:8–12. doi:10.1016/j.cimid.2016.11.005
104. Silva JR, Medeiros LC, Reis VP, et al. Serologic survey of West Nile virus in horses from Central-West, Northeast and Southeast Brazil. *Mem Inst Oswaldo Cruz*. 2013;108:921–923. doi:10.1590/S0074-02762013005000001
105. Prow NA, Tan CS, Wang W, et al. Natural exposure of horses to mosquito-borne flaviviruses in south-east Queensland, Australia. *Int J Environ Res Public Health*. 2013;10:4432–4443. doi:10.3390/ijerph10094432
106. Rocheleau JP, Michel P, Lindsay LR, et al. Emerging arboviruses in Quebec, Canada: assessing public health risk by serology in humans, horses and pet dogs. *Epidemiol Infect*. 2017;145:2940–2948. doi:10.1017/S0950268817002205

107. Cabre O, Grandadam M, Marie JL, et al. West Nile Virus in horses, sub-Saharan Africa. *Emerg Infect Dis*. 2006;12:1958–1960. doi:10.3201/eid1212.060042
108. Barbic L, Listes E, Katic S, et al. Spreading of West Nile virus infection in Croatia. *Vet Microbiol*. 2012;159:504–508. doi:10.1016/j.vetmic.2012.04.038
109. Durand B, Dauphin G, Zeller H, et al. Serosurvey for West Nile virus in horses in southern France. *Vet Rec*. 2005;157:711–713. doi:10.1136/vr.157.22.711
110. Bouzalas IG, Diakakis N, Chaintoutis SC, et al. Emergence of equine west nile encephalitis in central Macedonia, Greece, 2010. *Transbound Emerg Dis*. 2016;63:e219–e227. doi:10.1111/tbed.12334
111. Ludu Oslobanu EL, Mihu-Pintilie A, Anita D, Anita A, Lecollinet S, Savuta G. West Nile virus reemergence in Romania: a serologic survey in host species. *Vector Borne Zoonotic Dis*. 2014;14:330–337. doi:10.1089/vbz.2013.1405
112. Calistri P, Giovannini A, Savini G, et al. West Nile virus transmission in 2008 in north-eastern Italy. *Zoonoses Public Health*. 2010;57:211–219. doi:10.1111/j.1863-2378.2009.01303.x
113. Ibarra-Juarez L, Eisen L, Bolling BG, et al. Detection of West Nile virus-specific antibodies and nucleic acid in horses and mosquitoes, respectively, in Nuevo Leon State, northern Mexico, 2006-2007. *Med Vet Entomol*. 2012;26:351–354. doi:10.1111/j.1365-2915.2012.01014.x
114. Benjelloun A, El Harrak M, Calistri P, et al. Seroprevalence of West Nile virus in horses in different Moroccan regions. *Vet Med Sci*. 2017;3:198–207. doi:10.1002/vms3.71
115. Zohaib A, Saqib M, Beck C, et al. High prevalence of West Nile virus in equines from the two provinces of Pakistan. *Epidemiol Infect*. 2015;143:1931–1935. doi:10.1017/S0950268814002878
116. Bazanow B, Jansen van Vuren P, Szymanski P, et al. A survey on West Nile and Usutu viruses in horses and birds in Poland. *Viruses*. 2018;10:E87. doi:10.3390/v10020087
117. Hemida MG, Perera R, Chu DKW, Ko RLW, Alnaeem AA, Peiris M. West Nile virus infection in horses in Saudi Arabia (in 2013-2015). *Zoonoses Public Health*. 2019;66:248–253. doi:10.1111/zph.12532
118. Csank T, Drzewniokova P, Korytar L, et al. A SEROSURVEY OF FLAVIVIRUS INFECTION IN HORSES AND BIRDS IN Slovakia. *Vector Borne Zoonotic Dis*. 2018;18:206–213. doi:10.1089/vbz.2017.2216
119. Garcia-Bocanegra I, Arenas-Montes A, Napp S, et al. Seroprevalence and risk factors associated to West Nile virus in horses from Andalusia, Southern Spain. *Vet Microbiol*. 2012;160:341–346. doi:10.1016/j.vetmic.2012.06.027
120. Bargaoui R, Lecollinet S, Lancelot R. Mapping the serological prevalence rate of West Nile fever in equids, Tunisia. *Transbound Emerg Dis*. 2015;62:55–66. doi:10.1111/tbed.12077
121. Ergunay K, Gunay F, Erisoz Kasap O, et al. Serological, molecular and entomological surveillance demonstrates widespread circulation of West Nile virus in Turkey. *PLoS Negl Trop Dis*. 2014;8(7):e3028. doi:10.1371/journal.pntd.0003028
122. Ziegler U, Skrypnyk A, Keller M, et al. West nile virus antibody prevalence in horses of Ukraine. *Viruses*. 2013;5:2469–2482. doi:10.3390/v5102469
123. Franson JC, Hofmeister EK, Collins GH, Dusek RJ. Seroprevalence of West Nile virus in feral horses on Sheldon National Wildlife Refuge, Nevada, United States. *Am J Trop Med Hyg*. 2011;84:637–640. doi:10.4269/ajtmh.2011.10-0467
124. Bosch I, Herrera F, Navarro JC, et al. West Nile virus, Venezuela. *Emerg Infect Dis*. 2007;13:651–653. doi:10.3201/eid1304.061383

CHAPTER 3

Re-parameterisation of a mathematical model of African horse sickness virus using data from a systematic literature search

Note: This chapter is a published article:

Fairbanks EL, Brennan ML, Mertens PPC, Tildesley MJ, Daly JM. Re-parameterization of a mathematical model of African horse sickness virus using data from a systematic literature search. Transboundary and Emerging Diseases. 2021.

Supplementary material for this chapter can be found in Appendix E.

Author contribution: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing - review & editing.

Received: 20 July 2021 | Revised: 29 September 2021 | Accepted: 29 September 2021

DOI: 10.1111/tbed.14420

Transboundary and Emerging Diseases WILEY

ORIGINAL ARTICLE

Re-parameterization of a mathematical model of African horse sickness virus using data from a systematic literature search

Emma L. Fairbanks[1] | **Marnie L. Brennan**[1] | **Peter P. C. Mertens**[1] | **Michael J. Tildesley**[2] | **Janet M. Daly**[1]

[1] School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK

[2] The Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research, School of Life Sciences and Mathematics Institute, University of Warwick, Coventry, UK

Correspondence
Emma L. Fairbanks, School of Veterinary Medicine and Science, University of Nottingham, Nottingham, LE12 5RD, UK.
Email: emma.fairbanks@swisstph.ch

Funding information
BBSRC, Grant/Award Number: 1944283; EU H2020, Grant/Award Number: 727293 PALE-Blu

Abstract

African horse sickness (AHS) is a vector-borne disease transmitted by *Culicoides* spp., endemic to sub-Saharan Africa. There have been many examples of historic and recent outbreaks in the Middle East, Asia and Europe. However, not much is known about infection dynamics and outbreak potential in these naive populations. In order to better inform a previously published ordinary differential equation model, we performed a systematic literature search to identify studies documenting experimental infection of naive (control) equids in vaccination trials. Data on the time until the onset of viraemia, clinical signs and death after experimental infection of a naive equid and duration of viraemia were extracted. The time to viraemia was 4.6 days and the time to clinical signs was 4.9 days, longer than the previously estimated latent period of 3.7 days. The infectious periods of animals that died/were euthanized or survived were found to be 3.9 and 8.7 days, whereas previous estimations were 4.4 and 6 days, respectively. The case fatality was also found to be higher than previous estimations. The updated parameter values (along with other more recently published estimates from literature) resulted in an increase in the number of host deaths, decrease in the duration of the outbreak and greater prevalence in vectors.

KEYWORDS
African horse sickness, mathematical model, vector-borne disease

1 | INTRODUCTION

African horse sickness (AHS) is caused by African horse sickness virus (AHSV) of the genus *Orbivirus* in the family *Reoviridae*. Endemic to sub-Saharan Africa, it often emerges when periods of heavy rain follow hot and dry conditions. This favours its principle vector, *Culicoides* spp., with *Culicoides imicola* usually considered to be the most important vector species in Africa (Mellor & Hamblin, 2004). AHSV hosts include horses, mules, donkeys and zebras. Zebras, the only equids native to South Africa, are believed to be the reservoir host (Barnard, 1998).

There are nine immunologically distinct serotypes of AHSV defined on the basis of antigenic reactivity of antibodies to the outer capsid virus protein VP2 (Bachanek-Bankowska et al., 2014). In endemic regions, usually only one serotype circulates at a time. Historically, serotype 9 was responsible for epizootics of AHS outside Africa. Outbreaks in central and East Africa have occasionally spread to Egypt, the Middle East and southern Arabia (Mirchamsy & Hazrati, 1973). A major epidemic in 1959–1961, spread through the Near East and Arabia as far as Pakistan and India, resulted in the death of an estimated 300,000 equids (Anwar & Qureshi, 1973; Howell, 1960). A further epidemic of AHSV in northwest Africa (Morocco, Algeria and Tunisia) in 1965–1966 spread briefly to southern Spain but was eliminated by vaccination and by killing infected equids (Hazrati, 1967). The first occurrence of an outbreak outside Africa not caused by serotype 9 was

in July 1987, when serotype 4 was reported in central Spain (Lubroth, 1988). The virus overwintered and caused further outbreaks in southern Spain, Portugal and Morocco in subsequent years before it was eliminated in 1991 (Baylis et al., 1997; Portas et al., 1999; Rodriguez et al., 1992). In 2007, outbreaks of AHS in West Africa were caused by serotype 2 (Nigeria and Senegal) and serotype 7 (Senegal) (Diouf et al., 2013). Outbreaks of AHS have recently been reported in Ethiopia (Aklilu et al., 2014), Thailand (Lu et al., 2020) and Malaysia (Castillo-Olivares, 2021).

AHS presents as acute, subacute or subclinical forms (Carpenter et al., 2017). In naive populations of horses, case fatality may exceed 90% in epidemics (Castillo-Olivares, 2021). The acute respiratory form is characterized by a short incubation period (3–5 days), and the animal usually dies from severe hypoxia, congestive heart failure or a combination of both after around 1 week. The subacute or cardiac form of AHS has an incubation period of 1–2 weeks with a short fever followed by the classical clinical sign of AHS and oedema of the supraorbital fossae. The case fatality is around 50% with death usually occurring within 1 week. In most outbreaks, a mixed pulmonary and cardiac form is most commonly seen, which causes fatality of around 80% of susceptible equids (Theiler, 1921). The subclinical form, often referred to as African horse sickness fever, is common in zebras, African donkeys and horses that are partially immune because they have been vaccinated or have recovered from a previous infection (Lu et al., 2020). The outbreaks of AHSV in Spain in 1987 and Thailand in 2020 were associated with importation of (infected) zebras from Africa (Grewar et al., 2021; Rodriguez et al., 1992).

In order to model the risk posed by AHSV if it emerges in countries where equids have no prior exposure, we need data on various parameters such as the length of the latent and infectious periods. It is also important to know whether there is any association between the serotype causing an outbreak and the values of these parameters. In the case of AHSV, parameters such as the latent period are not possible to determine in the field as the exact time of infection is generally not known. Here, a systematic search and data extraction, focusing on studies documenting experimental infection of naive equids in vaccination trials, was performed to inform a model for AHSV previously suggested by Backer and Nodelijk (2011). Overall, 26 studies were used to derive parameters describing the host–virus interactions, compared to the three studies used in the development of the Backer and Nodelijk model (J. A. House et al., 1994; Roy et al., 1996; Scanlen et al., 2002). Parameters derived from the review were the time until the onset of viraemia, clinical signs and death after experimental infection of a naive equid as well as the duration of viraemia. It is important to consider the role of vectors when modelling AHSV. The parameters for the vectors were updated from literature, similarly to Gubbins et al. (2014); some of this literature was published after the Backer and Nodelijk model.

AHSV is listed as a notifiable disease in disease-free countries by the World Organisation for Animal Health (OIE) (2021). An outbreak can have a severe impact on the welfare of equids and be disruptive for the equine industry (Allison et al., 2009; Clemmons et al., 2021). Policy makers with responsibility for deciding the course of action if an outbreak occurs in a previously disease-free country are heavily reliant on models to predict the likely outcome of outbreaks (Daly et al., 2013; Grassly & Fraser, 2008).

2 | METHODS

2.1 | Systematic literature search

A systematic search (limited to the title and abstract) was performed (by E. L. Fairbanks and J. M. Daly) using PubMed with the search terms 'African horse sickness virus' AND ~vaccine AND (~challenge OR ~trial)' and 'African horse sickness virus' AND 'experimental infection', where ~ means terms including and similar too. Eligible studies included the inoculation of a naive equid with AHSV. The citations and references of eligible articles found on PubMed were then searched in order to find further relevant articles. A similar search term was used to search CAB abstracts; however, no additional research articles were identified. Supporting Information 1 describes how vector-related parameters were updated, also by systematic searches.

2.1.1 | Data extraction and analysis

Studies were eligible for data extraction if they gave a value for the time until onset of viraemia and/or clinical signs and/or death. To determine whether these attributes were significantly different between the virus serotypes, a Kruskal–Wallis test was used. This is a non-parametric method for testing whether samples originate from the same distribution, that can be used for comparing more than two independent samples. This analysis was performed using the stats R package version 4.2.0 (R Core Team and contributors worldwide, 2021). Data for the unidentified serotypes were not included in this analysis. Also, analysis for the time until viraemia was repeated including only serotypes with more than one data value. We also performed this statistical analysis to determine whether infection method (either intravenous or subcutaneous inoculation) had an impact on these disease characteristics. A paired and two-sample t-tests were used to determine if there was a difference between the estimated onset of viraemia between polymerase chain reaction (PCR)-based and virus isolation methods in the equids for which both were compared and all equids for which data for the onset of viraemia were analyzed (including both PCR and virus isolation for equines with both available), respectively.

2.2 | Re-parameterization of a mathematical model

We will consider a deterministic ordinary differential equation (ODE) model for AHSV suggested by Backer and Nodelijk (2011). In the Backer and Nodelijk model, the vector population adapts to changes in the vector:host ratio. Here, we do not include this in the model. This is due to the assumption that the size of the vector population is dependent on the carrying capacity of midges in the environment surrounding the horses rather than the number of horses present. The total number of vectors therefore does not change. We assume the number of hosts present is likely to reflect the size of the premises. Figure 1 shows

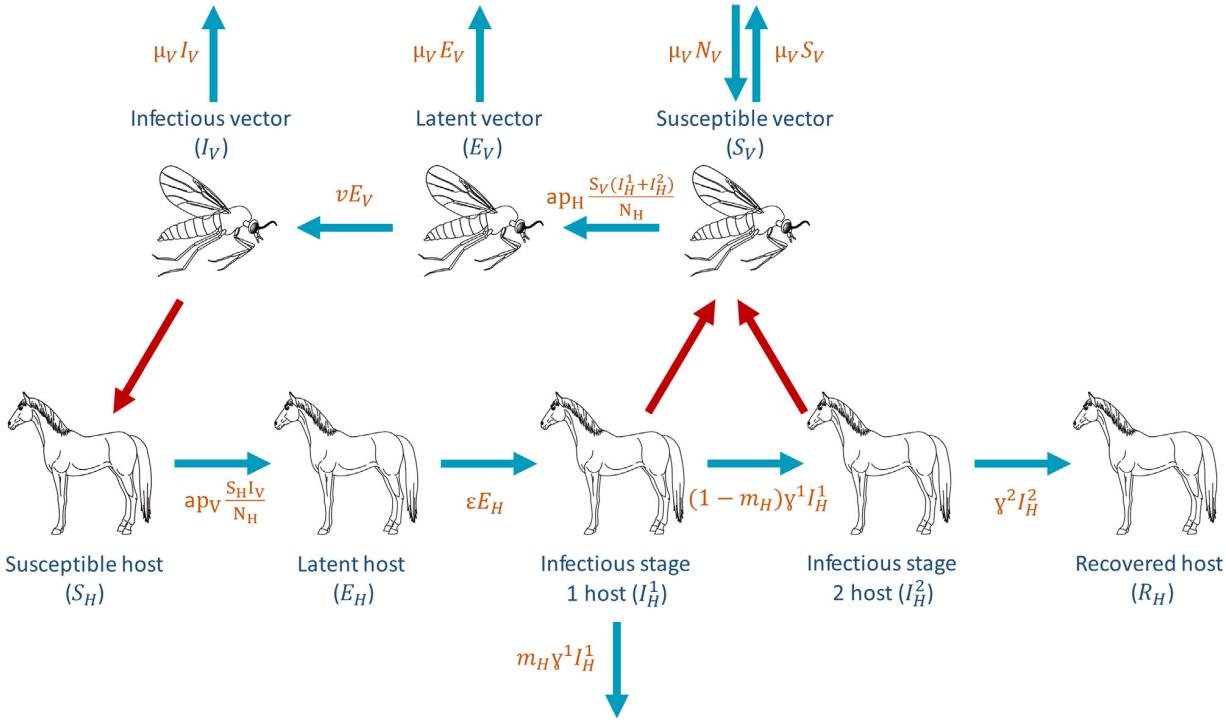

FIGURE 1 The infection cycle of African horse sickness virus (AHSV). Blue arrows represent transfers between compartments with their rates given in orange. Red arrows represent between species transmission

the infection diagram for this system. Details of the model along with the ODEs are given in Supporting Information 2.

Parameters updated included the latent period, infectious period of both dying and surviving equids and host fatality. The infectious period of surviving hosts was calculated using data from live virus isolation, similarly to Backer and Nodelijk (2011). Here, we set a conservative estimate of the case fatality of 0.84; however, sensitivity analysis allows us to observe the effect of this value on model outputs (Supporting Information 3). The latent period and infectious period of dying and surviving hosts are divided into multiple stages; this allows them to have a gamma distribution. For example, the total duration of the latent period follows a gamma distribution with mean duration ε and variance ε^2/i, where i is the number of stages.

We compare simulation results for this model for the parameters published in Backer and Nodelijk (2011) and the updated parameters. ODEs were integrated using the MATLAB function *ode45*. Sensitivity analysis was performed on the model with an updated range of possible parameter values. This was done on the duration of the outbreak, total number of infected equine hosts during the outbreak and the basic reproduction number (R_0) using the partial rank correlation coefficient (PRCC) method (Supporting Information 3).

3 | RESULTS

3.1 | Systematic literature search

A total of 39 articles were found during the PubMed search performed on 20 May 2020. Of these, 11 were excluded during a first screening of their titles for being narrative reviews or a non-equid experimental infection. After reading the full texts of the remaining 28 articles, a further 12 were excluded for being narrative reviews, in vitro experiments, non-equid infection experiments, not challenging equids or not using unvaccinated control animals, all of which mean that no naive equids were experimentally infected. The citations and references of the eligible 16 articles found on PubMed were then screened in order to find more relevant articles. A total of 40 additional full-text articles were then read and 13 were found to be eligible under the same inclusion criteria. During the screening, six titles were identified that qualified for a full text read but could not be accessed; these were not included in the qualitative synthesis. The process of the literature search is described by the PRISMA flow diagram in Figure 2 (Liberati et al., 2009).

Altogether, 29 studies were found recording the experimental infection of 61 naive equids; 53 horses, five donkeys and three mules (Alberca et al., 2014; Alexander & Du Toit, 1934; Du Plessis et al., 1998; Dubourget et al., 1992; El Hasnaoui et al., 1998; Guthrie et al., 2009; Hassanain, 1992; Hazrati & Ozawa, 1965; C. House et al., 1990; J. A. House et al., 1992, 1994; Lelli et al., 2013; Lulla et al., 2017; Martínez-Torrecuadrada et al., 1996, 1997; Minke et al., 2012; Mirchamsy & Taslimi, 1964a, 1964b, 1968; Ozawa & Bahrami, 1966; Ozawa et al., 1965, 1970; Quan et al., 2010; Roy et al., 1996; Scanlen et al., 2002; Stone-Marschat et al., 1996; van Rijn et al., 2018; von Teichman et al., 2010; Whitworth, 1930). These included all serotypes apart from AHSV-7 and two studies where the serotype was unidentified (Figure 3). The article by Sánchez-Matamoros et al. (2016) was considered, but not included due to uncertainty of the vaccination

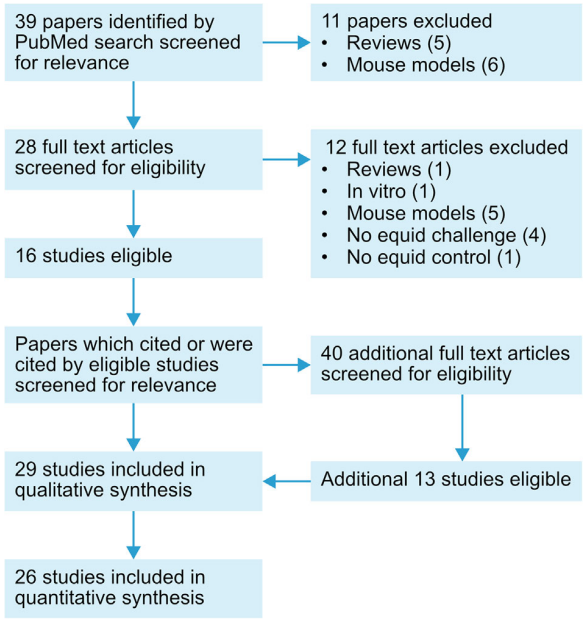

FIGURE 2 PRISMA flow diagram describing the process of the systematic search

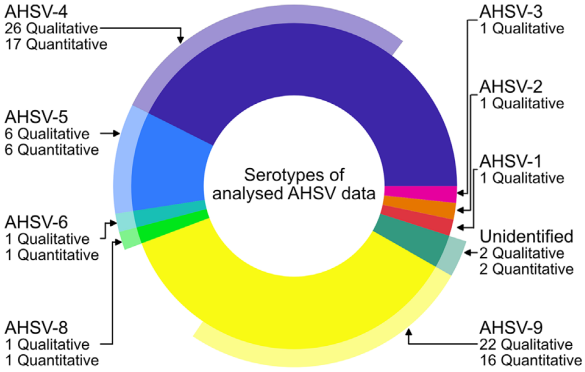

FIGURE 3 Pie chart showing the percentage of qualitative samples of each serotype from the systematic search (inner circle) and the percentage of these used in the quantitative analysis after data extraction (outer circle)

status of the 14 experimentally infected horses. Qualitative synthesis of the information available from papers can be found in Table S5. El Hasnaoui et al. (1998) was included in the qualitative analysis due to the naive experimental infections; however, the full text was not accessible so details about values of interest were not available.

3.1.1 | Data extraction and analysis

Of the 29 studies analyzed during the qualitative synthesis, 26 were eligible for data extraction and analysis (i.e. gave a value for the time until onset of viraemia and/or clinical signs and/or death).

A total of 27 naive experimental infections where the time until viraemia had been measured were analyzed (Figure 4a). These varied from 2 days to 11 days with mean 4.6 days and included serotypes AHSV-4 ($m = 14$), AHSV-5 ($m = 5$), AHSV-6 ($m = 1$), AHSV-8 ($m = 1$) and AHSV-9 ($m = 6$). The time until the onset of clinical signs was recorded for 21 of the naive experimental infections for serotypes AHSV-4 ($m = 9$), AHSV-5 ($m = 4$), AHSV-9 ($m = 6$) and two unidentified serotypes. The mean was 4.9 days and individuals varied from 3 days to 10 days (Figure 4b). The mean time until death after experimental infection was 9.1 days and varied between 5 and 17 days (Figure 4c). This was recorded for 25 horses for serotypes AHSV-4 ($m = 6$), AHSV-5 ($m = 5$), AHSV9 ($m = 13$) and an unidentified serotype ($m = 1$).

Results from the Kruskal–Wallis tests to determine whether the time to viraemia, onset of clinical signs or death are significantly different between the virus serotypes or inoculation methods are given in Table S6. None of these tests gave a significant result ($p < .05$); therefore, we conclude that there is no evidence that the time until viraemia, the onset of clinical signs or death varies between serotypes or inoculation method. A paired t-test showed that there was no significant difference between the start of viraemia between PCR and virus isolation methods in the six equids for which both were compared ($t = -1.55$, $df = 6$, $p = .17$). A two-sample t-test performed for the 27 equines for which data for the start of viraemia was analyzed also showed no statistically significant difference between detection of genetic material by PCR-based methods and detection of infectious virus ($t = 0.64$, $df = 31$, $p = .52$).

For 11 horses, the time until viraemia and death due to AHSV was recorded. The total infectious period (difference between onset of viraemia and death) of these horses varied between 2 and 7 days with a mean of 3.9 days (Figure 4d). For three horses that survived experimental infection, the time between the start and end of viraemia was recorded using virus isolation. The total infectious periods of these horses were 4, 5 and 5 days, giving a mean of 4.7 days.

The lag between the start of viraemia and onset of clinical signs was reported for 18 horses. This varied between −1 (horse showed clinical signs before viraemia) and 3 days, with a mean of 0.8 days (Figure 4e).

Whether the equid died of AHSV, was euthanized for ethical reasons or survived was reported for one donkey and 44 horses. The donkey survived the AHSV-9 experimental infection. Of the 44 horses 31 died, seven were euthanized and six survived. A further eight horses either died naturally due to infection or were euthanized, but this was not specified. If we assumed that the euthanized horses were determined to be ill enough that they would not survive then a total of six horses out of 52 survived and the case fatality is calculated to be 0.88. However, if we do not make this assumption and do not consider the seven horses that were euthanized, six horses out of 37 survived giving a case fatality of 0.84.

3.2 | Re-parameterization of a mathematical model

The model parameters updated from the review are given in Table 1. As well as these host parameters, some vector parameters were also

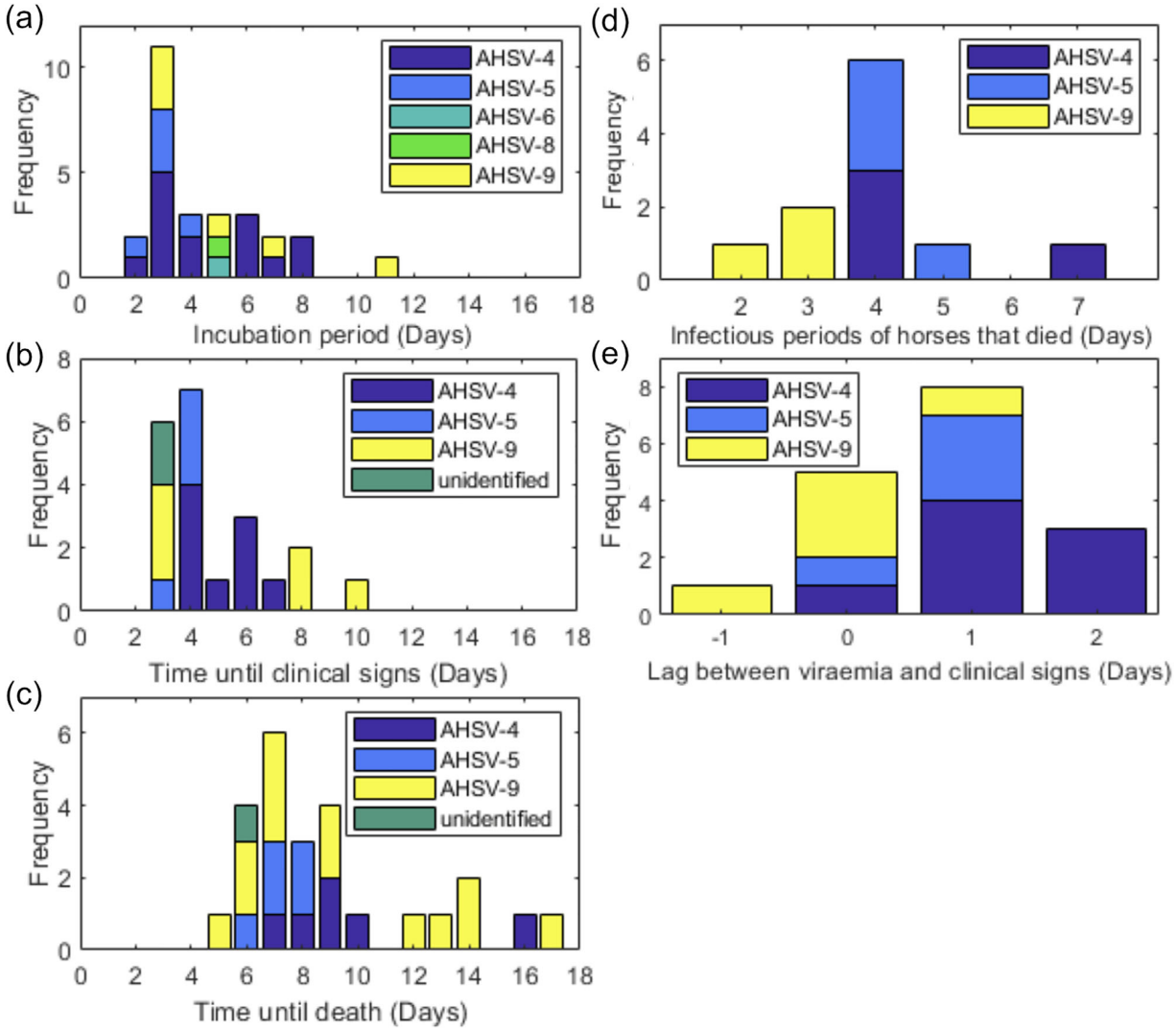

FIGURE 4 Histograms of (a) the time until viraemia, (b) time until clinical signs, (c) time until death, (d) total infectious period of horses that died and (e) time lag between viraemia and clinical signs for equids experimentally infected with African horse sickness virus (AHSV)

TABLE 1 Updated model parameters. Time periods are given in days. The parameter values used in Backer and Nodelijk (2011) are described in the previous value column and the parameter values used in our model are described in the updated value column

		Previous value		Updated value	
Parameter	Symbol	Default value	5%–95% range	Default value	5%–95% range
Latent period	$1/\varepsilon$	3.7	2.5–4.9	4.6	2–8.5
Number of stages	1	16		5	
Infectious period (dying hosts)	T_{inf}^{1}	4.4	2.2–6.6	3.9	2.1–6.9
Number of stages	n^1	19		11	
Infectious period (recovering hosts)	T_{inf}^{2}	6.0	3.0–9.0	4.7	4–5
Number of stages	n^2	10		13	
Host fatality	m_H	0.7	0.43–0.97	0.84–0.88	

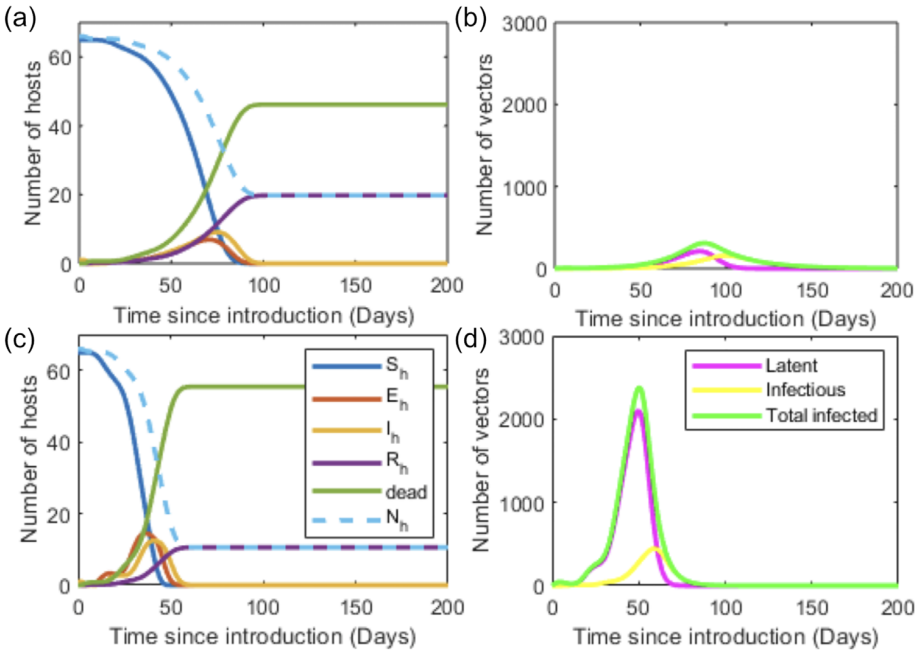

FIGURE 5 Simulations of the Backer and Nodelijk (2011) model using the published (a, b) and updated (c, d) parameter values showing the outcomes for the host (a, c) and vector (b, d) populations over time after the introduction of infection into a population of 66 horses. These parameter values are given in Table 1 and Table S1

updated from the literature. These were the blood feeding interval (frequency of biting) (Mullens et al., 2004), extrinsic incubation period (latent period of midges) (Carpenter et al., 2011), vector life-span (Gerry & Mullens, 2000) and transmission probability from a host to a vector during a bite (Carpenter et al., 2011). Details of these and other unchanged parameters are given in Supporting Information 1.

We fit a gamma distribution to the quantitative data for the time until viraemia and duration of viraemia to calculate the number of stages in the latent and infectious periods, respectively. The number of stages is equal to the gamma parameter. The number of stages for the latent period and infectious period for dying hosts are 5 (95% CI:3–9) and 11 (95% CI:5–26), respectively. Due to there not being many data points for the infectious period of surviving hosts, the number of stages needed was calculated as 93 (95% CI:19–659). This large value and confidence interval is due to the limited number of data points available. Therefore, due to the limited number of samples, the number of stages for the infectious period of surviving hosts is scaled from the infectious period of the dying hosts. As the mean infectious period of surviving hosts is 1.2 times longer than that of dying hosts, the infectious period of surviving hosts is divided into 13 stages (2 more than the number of stages for the dying horses).

Figure 5 compares the output of the updated parameters and the original Backer and Nodelijk model parameters detailed in Table 1 and Table S1. Initially, one horse is assumed to be in the first infectious stage (T_{inf}^1). The updated parameters suggest the outbreak would have a shorter duration, and the total number of host deaths increased from 46 to 55. For both sets of parameters, all hosts on the premises become infected. The updated parameter values suggest a peak of 13 infectious hosts on day 41 of the outbreak. However, the original parameter values suggest this peak is on day 75 with nine infectious hosts. The peak in the number of infectious vectors comes after this. The updated parameters suggest that the peak is 439 infectious vectors on day 58 of the outbreak compared to 156 infectious vectors on day 100 using the original parameters.

Sensitivity analysis (Supporting Information 3) showed the parameter that most significantly influences the duration of the outbreak is the vector:host ratio, with a larger vector population being associated with shorter outbreaks. Here, the host infectious period (T_{inf}), vector life-span ($1/\mu_V$), host to vector (p_H) and vector to host (p_V) transmission rates, and the vector:host ratio are associated with outbreaks that spread more rapidly upon emergence. It is also shown that longer host latent periods ($1/\varepsilon$), duration between vector bites ($1/a$) and extrinsic incubation periods ($1/\nu$) increase the duration of the outbreak. The host case fatality (m_H) did not significantly influence R_0.

4 | DISCUSSION

In order to better inform the host parameters used in the ODE model developed by Backer and Nodelijk (2011), data were extracted from 26 studies, representing 43 experimentally infected naive equids. Since this analysis, further naive experimental infection studies for AHSV have been published (van Rijn et al., 2020). These data did not conflict with the findings of this study. Overall, the updated model parameters suggest more hosts on the premises would die and the outbreak on each premises would be shorter than previously predicted. More

vectors would also become infected, which may increase the probability of transmission between premises.

Many of the studies in this systematic search use different volumes and titres of virus for experimental infection. Experimental infection through needle injection is not a natural route of infection. How well this mimics natural infection by midges is not known (Coetzee et al., 2014; Darpel et al., 2012). It is therefore unclear whether this route of infection or the titre/volume of virus inoculated affects the time taken to become viraemic or show clinical signs. Bluetongue virus experimental infections have shown that infection via the intradermal route in sheep reduced the time until clinical signs and increased their severity compared to the intravenous route (Umeshappa et al., 2011). It was also found that subcutaneous inoculation appears to mimic the natural route of infection more closely than the intravenous inoculation route, in respect to the dissemination of the virus from the skin to secondary target organs as has been observed following natural infection (Coetzee et al., 2014; Umeshappa et al., 2011). In this review, we found no difference in the time until viraemia, clinical signs or death between intravenous and subcutaneous routes of infection.

Variation in VP2 is a mechanism for escape from pre-existing neutralizing antibodies that block viral attachment to the host cell (Burrage et al., 1993; Jewell & Mecham, 1994). These variations may impact the infectivity of the virus for the cells of hosts or vectors, but viral pathogenicity is typically determined by more than one viral protein and is therefore not necessarily serotype dependent. For example, van Rijn et al. (2018) reported that introducing a deletion in the NS3/NS3a protein of AHSV renders the virus avirulent. Lulla et al. (2016) reported that an AHSV-1 isolate was more virulent than an AHSV-4 isolate in mice, but whether this difference was determined by VP2 or indeed consistent between different isolates of the respective serotypes was not addressed by that study. No differences in time until viraemia, clinical signs or death were found between serotypes in this study. Studies of bluetongue virus (BTV) (the type species of the *Reoviridae* family) indicate that more recently identified serotypes (BTV-25–27) show adaptations involving VP2 (as well as VP1, VP3 and VP7) that support direct-contact transmission, rather than by vector insects. These changes also influence the virulence of these serotypes for different ruminant hosts (Guimerà Busquets et al., 2021; Pullinger et al., 2016), but there is no evidence for similar variations in AHSV.

Articles analyzed to parameterize this model dated back as far as 1930, that is, before PCR-based techniques were available. In one study, the reverse transcription-PCR (RT-PCR) assay and virus isolation methods were found to be equally sensitive for detection of virus in blood samples from horses experimentally infected with AHSV-4. However, viraemia was detected more consistently by RT-PCR than by virus isolation from horses infected with AHSV-9 except from one animal for which virus was detected only by virus isolation (Sailleau et al., 1997). Other studies have shown that RT-quantitative PCR has higher sensitivity than virus isolation (Guthrie et al., 2013; Quan et al., 2010). Three articles, with a total of six experimental infections of naive equids from which data were extracted in this study, compared the detection of viraemia by PCR and virus isolation methods. Of these, two studies (involving a total of four equids) (Alberca et al., 2014; Guthrie et al., 2009) did not find a difference in the first day of detection of viraemia, whereas one found that viraemia was detected one day earlier by PCR than by virus isolation in both animals involved (Lelli et al., 2013). Further analysis showed no significant differences between virus detection methods and estimated onset of viraema. However, caution is needed when interpreting RT-PCR results in relation to the duration of viraemia. For example, in BTV infection, nucleic acid can be detected in blood of hosts after infectious virus has been cleared (MacLachlan, 1994; MacLachlan et al., 1994; Mayo et al., 2021). Consequently, a long duration of viraemia detected by PCR is not a reliable guide to infectivity, or the ability of the host to act as a source of virus to infect feeding insects. The surviving host will have developed neutralizing antibodies that effectively inhibit detection by virus isolation, even if the virus was still viable (which is uncertain).

Many vector-borne disease models consider daily differences in temperature and seasonality in the vector population. Here, we do not consider seasonality; we simulate the model for the time of year at which midge-borne diseases are thought to be most likely to emerge in the United Kingdom. The midge species and climates in laboratory/field experiments used to derive the midge parameters may not be an accurate representation of the most likely vector in all regions where there is a risk of AHSV. The use of the Gerry and Mullens (2000) value for midge mortality compared to the Wittmann et al. (2002) value reduces estimates for transmission. A limitation of Wittmann et al. (2002) is that *Culicoides* were examined in the laboratory which may affect their mortality in comparison to wild *Culicoides*. However, the parameter estimated from Gerry and Mullens (2000), based on trap data, also has its limitations. A recent study found that biting rate per day (per cow) was expected to be approximately 50% of a 24-h trap collection (Möhlmann et al., 2021). Another study found that on average 2.2 times more *Culicoides* were found on a cow than on a Shetland pony (Elbers & Meiswinkel, 2015). However, this could be variable between species and size of horses/cows. There is a lot of uncertainty around this parameter; therefore, the range of values considered during the sensitivity analysis was large.

Some parameters such as the length of the host and vector latent periods ($1/\varepsilon$ and $1/\nu$, respectively) and the case fatality (m_H) cannot be influenced by control measures. To reduce the infectious period of hosts, infected hosts are often euthanized. Analysis of the studies from the systematic search showed that clinical signs appear a mean of 0.8 days after onset of viraemia. Euthanasia would only take place after clinical signs are observed (which is likely to be very variable), a vet responds, and makes a diagnosis. Therefore, at the time of euthanasia, a proportion of the infectious period is likely to have already passed. Results show that the length of the infectious period does not influence the duration of the outbreak as much as the vector:host ratio. This suggests that vector control may be a more efficient method of controlling outbreaks. Vector control reduces the life-span of the midges ($1/\mu_V$) and therefore the probability they complete the extrinsic incubation period and become infectious is also reduced. There are several methods of reducing the bite rate of vectors ($1/a$); these include insect repellents, fly rugs, stabling horses during times of high midge activity and insect-proofing stable areas (Barba et al., 2019; Meiswinkel et al., 2000;

Robin et al., 2015). Vaccination to prevent the horse from becoming infected reduces (to 0 in a perfect vaccine) the transmission probability from vector to host (p_V). The transmission probability from hosts to vectors (p_H) would also likely be reduced if a partially immune animal became infected due to lower levels of viraemia (Scanlen et al., 2002).

Euthanasia of infected horses for ethical reasons affects our ability to assess the case fatality. However, as mentioned previously, during an outbreak it is likely that equids with severe clinical signs that are unlikely to survive would be euthanized. Therefore, including these euthanized hosts when considering those that do not survive would not be unrealistic in the model. The longer infectious period of surviving hosts may also be informative for policy makers when deciding if culling should be used as a control strategy. The rapid onset of death in AHSV-infected naive horses is itself thought to limit virus spread. The persistence (long term) of AHSV in the field is thought to be heavily dependent on the availability and distribution of alternative, possibly asymptomatic hosts, such as zebras or potentially donkeys or mules. Therefore, it is important to consider secondary hosts and their ability to aid spread of AHSV between premises, even if they are not specifically included in the model.

Overall, the re-parameterization of the model is more informed by the literature. Robin et al. (2016) stated that 'extensive further research is required if the equine industry is to avoid or effectively contain an AHS epizootic in disease-free regions'. Here, four key areas for further research were highlighted. These included improving the accuracy of disease modelling, which we have aimed to address in this study. This study also supports the importance of further research on the vector competence of certain *Culicoides* species and our knowledge of their distribution due to the lack of robust literature to parameterize these aspects of the model. Other areas highlighted included methods of reducing transmission, such as vaccination and methods of reducing vector bite rates. Sensitivity analysis highlighted the importance of vector control, supporting the suggestion in Robin et al. (2016) that developing more effective and practical methods of preventing *Culicoides* blood-feeding on horses may be key in AHSV control.

ACKNOWLEDGEMENTS

Emma L. Fairbanks acknowledges support from BBSRC award ref. 1944283. Peter P. C. Mertens acknowledges support from EU H2020 grant number 727293 PALE-Blu.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICS STATEMENT

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required.

DATA AVAILABILITY STATEMENT

All data used in this manuscript are available from the references cited.

ORCID

Emma L. Fairbanks https://orcid.org/0000-0002-1598-962X

Janet M. Daly https://orcid.org/0000-0002-1912-4500

REFERENCES

Aklilu, N., Batten, C., Gelaye, E., Jenberie, S., Ayelet, G., Wilson, A., Belay, A., Asfaw, Y., Oura, C., Maan, S., Bachanek-Bankowska, K., & Mertens, P. P. C. (2014). African horse sickness outbreaks caused by multiple virus types in Ethiopia. *Transboundary and Emerging Diseases*, *61*(2), 185–192.

Alberca, B., Bachanek-Bankowska, K., Cabana, M., Calvo-Pinilla, E., Viaplana, E., Frost, L., Gubbins, S., Urniza, A., Mertens, P., & Castillo-Olivares, J. (2014). Vaccination of horses with a recombinant modified vaccinia Ankara virus (MVA) expressing African horse sickness (AHS) virus major capsid protein VP2 provides complete clinical protection against challenge. *Vaccine*, *32*(29), 3670–3674. https://doi.org/10.1016/j.vaccine.2014.04.036

Alexander, R. A., & Du Toit, P. J. (1934). The immunization of horses and mules against horse sickness by means of the neurotropic virus of mice and guinea pigs. *Onderstepoort Journal of Veterinary Research*, *2*, 375–391.

Allison, K., Taylor, N., Upton, M., & Wilsmore, T. (2009). African horse sickness. *Impact on the UK Horse Industry*. http://www.veeru.reading.ac.uk/documents/AHS

Anwar, M., & Qureshi, M. (1973). Control and eradication of African horse sickness in Pakistan. *Developmental Biology*, *199*, 225–228.

Bachanek-Bankowska, K., Maan, S., Castillo-Olivares, J., Manning, N. M., Maan, N. S., Potgieter, A. C., Di Nardo, A., Sutton, G., Batten, C., & Mertens, P. P. C. (2014). Real time RT-PCR assays for detection and typing of African horse sickness virus. *Plos One*, *9*(4), e93758.

Backer, J. A., & Nodelijk, G. (2011). Transmission and control of African horse sickness in The Netherlands: A model analysis. *Plos One*, *6*(8), e23066. https://doi.org/10.1371/journal.pone.0023066

Barba, M., Fairbanks, E. L., & Daly, J. (2019). Equine viral encephalitis: Prevalence, impact, and management strategies. *Veterinary Medicine*, *10*, 99–110. https://doi.org/10.2147/VMRR.S168227

Barnard, B. (1998). Epidemiology of African horse sickness and the role of the zebra in South Africa. *African Horse Sickness*, 13–19. https://doi.org/10.1007/978-3-7091-6823-3\underline{}2

Baylis, M., Hasnaoui, H. E., Bouayoune, H., Touti, J., & Mellor, P. S. (1997). The spatial and seasonal distribution of African horse sickness and its potential Culicoides vectors in Morocco. *Medical and Veterinary Entomology*, *11*(3), 203–212. https://doi.org/10.1111/j.1365-2915.1997.tb00397.x

Burrage, T. G., Trevejo, R., Stone-Marschat, M., & Laegreid, W. W. (1993). Neutralizing epitopes of African horsesickness virus serotype 4 are located on VP2. *Virology*, *196*(2), 799–803. https://doi.org/10.1006/viro.1993.1537

Carpenter, S., Mellor, P. S., Fall, A. G., Garros, C., & Venter, G. J. (2017). African horse sickness virus: History, transmission, and current status. *Annual Review of Entomology*, *62*, 343–358. https://doi.org/10.1146/annurev-ento-031616-035010

Carpenter, S., Wilson, A., Barber, J., Veronesi, E., Mellor, P., Venter, G., & Gubbins, S. (2011). Temperature dependence of the extrinsic incubation period of orbiviruses in Culicoides biting midges. *Plos One*, *6*(11), e27987. https://doi.org/10.1371/journal.pone.0027987

Castillo-Olivares, J. (2021). African horse sickness in Thailand: Challenges of controlling an outbreak by vaccination. *Equine Veterinary Journal*, *53*(1), 9–14. https://doi.org/10.1111/evj.13353

Clemmons, E. A., Alfson, K. J., & Dutton, J. W. (2021). Transboundary animal diseases, an overview of 17 diseases with potential for global spread and serious consequences. *Animals*, *11*(7), 2039. https://doi.org/10.3390/ani11072039

Coetzee, P., Van Vuuren, M., Venter, E. H., & Stokstad, M. (2014). A review of experimental infections with bluetongue virus in the mammalian host. *Virus Research*, *182*, 21–34. https://doi.org/10.1016/j.virusres.2013.12.044

Daly, J. M., Newton, J. R., Wood, J. L. N., & Park, A. W. (2013). What can mathematical models bring to the control of equine influenza? *Equine Veterinary Journal*, *45*(6), 784–788. https://doi.org/10.1111/evj.12104

Darpel, K. E., Monaghan, P., Simpson, J., Anthony, S. J., Veronesi, E., Brooks, H. W., Elliott, H., Brownlie, J., Takamatsu, H.-H., Mellor, P. S., & Mertens, P. P. (2012). Involvement of the skin during bluetongue virus infection and replication in the ruminant host. *Veterinary Research*, *43*(1), 40. https://doi.org/10.1186/1297-9716-43-40

Diouf, N. D., Etter, E., Lo, M. M., Lo, M., & Akakpo, A. J. (2013). Outbreaks of African horse sickness in Senegal, and methods of control of the 2007 epidemic. *Veterinary Record*, *172*(6), 152. https://doi.org/10.1136/vr.101083

Dubourget, P., Preaud, J., Detraz, N., Lacoste, F., Fabry, A., Erasmus, B., & Lombard, M. (1992). Development, production and quality control of an industrial inactivated vaccine against African horse sickness virus serotype 4. *Bluetongue, African horse sickness and related orbiviruses* (pp. 874–886). CRC Press.

Du Plessis, M., Cloete, M., Aitchison, H., & Van Dijk, A. (1998). Protein aggregation complicates the development of baculovirus-expressed African horse sickness virus serotype 5 VP2 subunit vaccines. *Onderstepoort Journal of Veterinary Research*, *65*(4), 321–329.

Elbers, A. R. W., & Meiswinkel, R. (2015). Culicoides (Diptera: Ceratopogonidae) and livestock in the Netherlands: Comparing host preference and attack rates on a Shetland pony, a dairy cow, and a sheep. *Journal of Vector Ecology*, *40*(2), 308–317. https://doi.org/10.1111/jvec.12169

El Hasnaoui, H., El Harrak, M., Zientara, S., Laviada, M., & Hamblin, C. (1998). Serological and virological responses in mules and donkeys following inoculation with African horse sickness virus serotype 4. *Archives of Virology Supplementum*, *14*, 29–36.

Gerry, A. C., & Mullens, B. A. (2000). Seasonal abundance and survivorship of *Culicoides sonorensis* (Diptera: Ceratopogonidae) at a southern California dairy, with reference to potential bluetongue virus transmission and persistence. *Journal of Medical Entomology*, *37*(5), 675–688. https://doi.org/10.1603/0022-2585-37.5.675

Grassly, N. C., & Fraser, C. (2008). Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, *6*(6), 477–487. https://doi.org/10.1038/nrmicro1845

Grewar, J. D., Kotze, J. L., Parker, B. J., Van Helden, L. S., & Weyer, C. T. (2021). An entry risk assessment of African horse sickness virus into the controlled area of South Africa through the legal movement of equids. *Plos One*, *16*(5), e0252117. https://doi.org/10.1371/journal.pone.0252117

Gubbins, S., Turner, J., Baylis, M., Van Der Stede, Y., Van Schaik, G., Abrahantes, J. C., & Wilson, A. J. (2014). Inferences about the transmission of Schmallenberg virus within and between farms. *Preventive Veterinary Medicine*, *116*(4), 380–390. https://doi.org/10.1016/j.prevetmed.2014.04.011

Guimerà Busquets, M., Pullinger, G. D., Darpel, K. E., Cooke, L., Armstrong, S., Simpson, J., Palmarini, M., Fragkoudis, R., & Mertens, P. P. C. (2021). An early block in the replication of the atypical bluetongue virus serotype 26 in culicoides cells is determined by its capsid proteins. *Viruses*, *13*(5), 919. https://doi.org/10.3390/v13050919

Guthrie, A. J., Maclachlan, N. J., Joone, C., Lourens, C. W., Weyer, C. T., Quan, M., Monyai, M. S., & Gardner, I. A. (2013). Diagnostic accuracy of a duplex real-time reverse transcription quantitative PCR assay for detection of African horse sickness virus. *Journal of Virological Methods*, *189*(1), 30–35. https://doi.org/10.1016/j.jviromet.2012.12.014

Guthrie, A. J., Quan, M., Lourens, C. W., Audonnet, J.-C., Minke, J. M., Yao, J., He, L., Nordgren, R., Gardner, I. A., & Maclachlan, N. J. (2009). Protective immunization of horses with a recombinant canarypox virus vectored vaccine co-expressing genes encoding the outer capsid proteins of African horse sickness virus. *Vaccine*, *27*(33), 4434–4438. https://doi.org/10.1016/j.vaccine.2009.05.044

Hassanain, M. M. (1992). Preliminary findings for an inactivated African horse sickness vaccine using binary ethyleneimine. *Revue D Elevage Et De Medecine Veterinaire Des Pays Tropicaux*, *45*(3–4), 231–234.

Hazrati, A. (1967). Identification and typing of horse-sickness virus strains isolated in the recent epizootic of the disease in Morocco, Tunisia and Algeria. *Archives of Razi Institute*, *19*(1), 131–143. https://doi.org/10.22092/ari.1967.108632

Hazrati, A., & Ozawa, Y. (1965). Monovalent live-virus horse-sickness vaccine. *Bulletin - Office International Des Epizooties*, *64*, 683–695.

House, C., Mikiciuk, P. E., & Beminger, M. L. (1990). Laboratory diagnosis of African horse sickness: Comparison of serological techniques and evaluation of storage methods of samples for virus isolation. *Journal of Veterinary Diagnostic Investigation*, *2*(1), 44–50. https://doi.org/10.1177/104063879000200108

House, J. A., Lombard, M., Dubourget, P., House, C., & Mebus, C. A. (1994). Further studies on the efficacy of an inactivated African horse sickness serotype 4 vaccine. *Vaccine*, *12*(2), 142–144. https://doi.org/10.1016/0264-410x(94)90052-3

House, J. A., Lombard, M., House, C., Dubourget, P., & Mebus, C. A. (1992). Efficacy of an inactivated vaccine for African horse sickness serotype 4. *Bluetongue, African horse sickness and related orbiviruses* (pp. 891–895). CRC Press.

Howell, P. (1960). The 1960 epizootic of African horse sickness in the Middle East and SW Asia. *Journal of the South African Veterinary Association*, *31*(3), 329–334. https://doi.org/10.10520/AJA00382809_817

Jewell, J. E., & Mecham, J. O. (1994). Identification of an amino acid on VP2 that affects neutralization of bluetongue virus serotype 10. *Virus Research*, *33*(2), 139–144. https://doi.org/10.1016/0168-1702(94)90050-7

Lelli, R., Molini, U., Ronchi, G. F., Rossi, E., Franchi, P., Ulisse, S., Armillotta G. , Capista S. , Khaiseb S. , Ventura M. D. , & Pini A. (2013). Inactivated and adjuvanted vaccine for the control of the African horse sickness virus serotype 9 infection: Evaluation of efficacy in horses and guinea-pig model. *Veterinaria Italiana*, *49*(1), 89–98.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *BMJ (Clinical Research Ed.)*, *339*, b2700. https://doi.org/10.1136/bmj.b2700

Lu, G., Pan, J., Ou, J., Shao, R., Hu, X., Wang, C., & Li, S. (2020). African horse sickness: Its emergence in Thailand and potential threat to other Asian countries. *Transboundary and Emerging Diseases*, *67*(5), 1751–1753. https://doi.org/10.1111/tbed.13625

Lubroth, J. (1988). African horsesickness and the epizootic in Spain 1987. *Equine Practice (USA)*, *10*(2), 26–33.

Lulla, V., Losada, A., Lecollinet, S., Kerviel, A., Lilin, T., Sailleau, C., Beck, C., Zientara, S., & Roy, P. (2017). Protective efficacy of multivalent replication-abortive vaccine strains in horses against African horse sickness virus challenge. *Vaccine*, *35*(33), 4262–4269. https://doi.org/10.1016/j.vaccine.2017.06.023

Lulla, V., Lulla, A., Wernike, K., Aebischer, A., Beer, M., & Roy, P. (2016). Assembly of replication-incompetent African horse sickness virus particles: Rational design of vaccines for all serotypes. *Journal of Virology*, *90*(16), 7405–7414. https://doi.org/10.1128/JVI.00548-16

Maclachlan, N. J. (1994). The pathogenesis and immunology of bluetongue virus infection of ruminants. *Comparative Immunology, Microbiology and Infectious Diseases*, *17*(3–4), 197–206. https://doi.org/10.1016/0147-9571(94)90043-4

Maclachlan, N. J., Nunamaker, R. A., Katz, J. B., Sawyer, M. M., Akita, G. Y., Osburn, B. I., & Tabachnick, W. J. (1994). Detection of bluetongue virus in the blood of inoculated calves: Comparison of virus isolation, PCR assay, and *in vitro* feeding of *Culicoides variipennis*. *Archives of Virology*, *136*(1–2), 1–8. https://doi.org/10.1007/BF01538812

Martinez-Torrecuadrada, J. L., Diaz-Laviada, M., Roy, P., Sanchez, C., Vela, C., Sanchez-Vizcaino, J. M., & Casal, J. I. (1996). Full protection against African horsesickness (AHS) in horses induced by baculovirus-derived AHS virus serotype 4 VP2, VP5 and VP7. *Journal of General Virology*, *77*(6), 1211–1221. https://doi.org/10.1099/0022-1317-77-6-1211

Martínez-Torrecuadrada, J. L., Díaz-Laviada, M., Roy, P., Sánchez, C., Vela, C., Sánchez-Vizcaíno, J. M., & Casal, J. I. (1997). Serologic markers in early stages of African horse sickness virus infection. *Journal of Clinical Microbiology*, *35*(2), 531–535.

Mayo, C. E., Weyer, C. T., Carpenter, M. J., Reed, K. J., Rodgers, C. P., Lovett, K. M., Guthrie, A. J., Mullens, B. A., Barker, C. M., Reisen, W. K., & Maclachlan, N. J. (2021). Diagnostic applications of molecular and serological assays for bluetongue and African horse sickness. *Revue Scientifique et Technique (International Office of Epizootics)*, *40*(1), 91–104. https://doi.org/10.20506/rst.40.1.3210

Meiswinkel, R., Baylis, M., & Labuschagne, K. (2000). Stabling and the protection of horses from *Culicoides bolitinos* (Diptera: Ceratopogonidae), a recently identified vector of African horse sickness. *Bulletin of Entomological Research*, *90*(6), 509–515. https://doi.org/10.1017/S0007485300000626

Mellor, P. S., & Hamblin, C. (2004). African horse sickness. *Veterinary Research*, *35*(4), 445–466. https://doi.org/10.1051/vetres:2004021

Minke, J. M., Audonnet, J.-C., Guthrie, A. J., MacLachlan, N. J., & Yao, J. (2012). *Vaccine against African horse sickness virus*. (U.S. Patent No. 8,168,200). Google Patents.

Mirchamsy, H., & Hazrati, A. (1973). A review on aetiology and pathogeny of African horse sickness. *Archives of Razi Institute*, *25*(1), 23–46. https://doi.org/10.22092/ARI.1973.108746

Mirchamsy, H., & Taslimi, H. (1964a). Attempts to vaccinate foals with living tissue culture adapted horse sickness virus. *Arch Inst Razi*, *17*(1), 17–27. https://doi.org/10.22092/ari.1965.108547

Mirchamsy, H., & Taslimi, H. (1964b). Immunization against African horse-sickness with tissue culture adapted neurotropic viruses. *British veterinary Journal*, *120*(10), 481–486. https://doi.org/10.1016/S0007-1935(17)41556-9

Mirchamsy, H., & Taslimi, H. (1968). Inactivated African horse sickness virus cell culture vaccine. *Immunology*, *14*(1), 81–88.

Möhlmann, T. W. R., Keeling, M. J., Wennergren, U., Favia, G., Santman-Berends, I., Takken, W., Koenraadt, C. J. M., & Brand, S. P. C. (2021). Biting midge dynamics and bluetongue transmission: A multiscale model linking catch data with climate and disease outbreaks. *Scientific Reports*, *11*(1), 1–16. https://doi.org/10.1038/s41598-021-81096-9

Mullens, B., Gerry, A., Lysyk, T., & Schmidtmann, E. (2004). Environmental effects on vector competence and virogenesis of bluetongue virus in Culicoides: Interpreting laboratory data in a field context. *Veterinaria Italiana*, *40*(3), 160–166.

Ozawa, Y., & Bahrami, S. (1966). African horse-sickness killed-virus tissue culture vaccine. *Canadian Journal of Comparative Medicine and Veterinary Science*, *30*(11), 311–314.

Ozawa, Y., Hazrati, A., & Bahrami, S. (1970). African horse-sickness live and killed virus tissue culture vaccine. *Archives of Razi Institute*, *22*, 103–111.

Ozawa, Y., Hazrati, A., & Erol, N. (1965). African horse-sickness live-virus tissue culture vaccine. *Archives of Razi Institute*, *18*, 61–84.

Portas, M., Boinas, F. S., Sousa, J. O. E., & Rawlings, P. (1999). African horse sickness in Portugal: A successful eradication programme. *Epidemiology and Infection*, *123*(2), 337–346. https://doi.org/10.1017/S0950268899002897

Pullinger, G. D., Guimerà Busquets, M., Nomikou, K., Boyce, M., Attoui, H., & Mertens, P. P. (2016). Identification of the genome segments of bluetongue virus serotype 26 (isolate kuw2010/02) that restrict replication in a *Culicoides sonorensis* cell line (kc cells). *Plos One*, *11*(2), e0149709. https://doi.org/10.1371/journal.pone.0149709

Quan, M., Lourens, C. W., Maclachlan, N. J., Gardner, I. A., & Guthrie, A. J. (2010). Development and optimisation of a duplex real-time reverse transcription quantitative PCR assay targeting the VP7 and NS2 genes of African horse sickness virus. *Journal of Virological Methods*, *167*(1), 45–52. https://doi.org/10.1016/j.jviromet.2010.03.009

R Core Team and contributors worldwide. (2021). Stats: The R Stats Package version 4.2.0 [Computer software manual]. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

Robin, M., Archer, D., Mcgowan, C., Garros, C., Gardès, L., & Baylis, M. (2015). Repellent effect of topical deltamethrin on blood feeding by Culicoides on horses. *Veterinary Record*, *176*(22), 574. https://doi.org/10.1136/vr.102800

Robin, M., Page, P., Archer, D., & Baylis, M. (2016). African horse sickness: The potential for an outbreak in disease-free regions and current disease control and elimination techniques. *Equine Veterinary Journal*, *48*(5), 659–669. https://doi.org/10.1111/evj.12600

Rodriguez, M., Hooghuis, H., & Castaño, M.a. (1992). African horse sickness in Spain. *Veterinary Microbiology*, *33*(1–4), 129–142. https://doi.org/10.1016/0378-1135(92)90041-Q

Roy, P., Bishop, D. H. L., Howard, S., Aitchison, H., & Erasmus, B. (1996). Recombinant baculovirus-synthesized African horse sickness virus (AHSV) outer-capsid protein VP2 provides protection against virulent AHSV challenge. *Journal of General Virology*, *77*(9), 2053–2057. https://doi.org/10.1099/0022-1317-77-9-2053

Sailleau, C., Moulay, S., Cruciere, C., Laegreid, W. W., & Zientara, S. (1997). Detection of African horse sickness virus in the blood of experimentally infected horses: Comparison of virus isolation and a PCR assay. *Research in Veterinary Science*, *62*(3), 229–232. https://doi.org/10.1016/S0034-5288(97)90195-8

Sánchez-Matamoros, A., Nieto-Pelegrín, E., Beck, C., Rivera-Arroyo, B., Lecollinet, S., Sailleau, C., Zientara, S., & Sánchez-Vizcaíno, J. M. (2016). Development of a luminex-based DIVA assay for serological detection of african horse sickness virus in horses. *Transboundary and Emerging Diseases*, *63*(4), 353–359. https://doi.org/10.1111/tbed.12503

Scanlen, M., Paweska, J. T., Verschoor, J. A., & Van Dijk, A. A. (2002). The protective efficacy of a recombinant VP2-based African horse sickness subunit vaccine candidate is determined by adjuvant. *Vaccine*, *20*(7-8), 1079–1088. https://doi.org/10.1016/s0264-410x(01)00445-5

Stone-Marschat, M. A., Moss, S. R., Burrage, T. G., Barber, M. L., Roy, P., & Laegreid, W. W. (1996). Immunization with VP2 is sufficient for protection against lethal challenge with African horse sickness virus Type 4. *Virology*, *220*(1), 219–222. https://doi.org/10.1006/viro.1996.0304

Van Rijn, P. A., Maris-Veldhuis, M. A., Grobler, M., Wright, I. M., Erasmus, B. J., Maartens, L. H., & Potgieter, C. A. (2020). Safety and efficacy of inactivated African horse sickness (AHS) vaccine formulated with different adjuvants. *Vaccine*, *38*(45), 7108–7117. https://doi.org/10.1016/j.vaccine.2020.08.072

Van Rijn, P. A., Maris-Veldhuis, M. A., Potgieter, C. A., & Van Gennip, R. G. P. (2018). African horse sickness virus (AHSV) with a deletion of 77 amino acids in NS3/NS3a protein is not virulent and a safe promising AHS Disabled Infectious Single Animal (DISA) vaccine platform. *Vaccine*, *36*(15), 1925–1933. https://doi.org/10.1016/j.vaccine.2018.03.003

Theiler, A. (1921). African horse sickness. *Science Bulletin*, *19*, 131–143. https://doi.org/10.22092/ari.1967.108632

Umeshappa, C. S., Singh, K. P., Channappanavar, R., Sharma, K., Nanjundappa, R. H., Saxena, M., Singh, R., & Sharma, A. K. (2011). A comparison of intradermal and intravenous inoculation of bluetongue virus serotype 23 in sheep for clinico-pathology, and viral and immune responses. *Veterinary Immunology and Immunopathology*, *141*(3–4), 230–238. https://doi.org/10.1016/j.vetimm.2011.03.005

Von Teichman, B. F., Dungu, B., & Smit, T. K. (2010). In vivo cross-protection to African horse sickness Serotypes 5 and 9 after vaccination with Serotypes 8 and 6. *Vaccine*, *28*(39), 6505–6517. https://doi.org/10.1016/j.vaccine.2010.06.105

Whitworth, S. (1930). Memorandum on horse-sickness immunization. In *Papers veterinary section, no. 30*. Union of South Africa, Dept. of Agriculture; Pretoria: University of Pretoria, Dept. of Library Services (Digital publisher). https://repository.up.ac.za/bitstream/handle/2263/13834/30_paper30_whitworth.pdf?sequence=1

Wittmann, E. J., Mellor, P. S., & Baylis, M. (2002). Effect of temperature on the transmission of orbiviruses by the biting midge, *Culicoides sonorensis*. *Medical and Veterinary Entomology*, *16*(2), 147–156. https://doi.org/10.1046/j.1365-2915.2002.00357.x

World Organisation for Animal Health (OIE). (2021). *Infection with African horse sickness virus - article 12.1.2*. https://www.oie.int/en/what-we-do/standards/codes-and-manuals/terrestrial-code-online-access/?id=169&L=1&htmfile=chapitre_ahs.htm

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Fairbanks, E. L., Brennan, M. L., Mertens, P. P. C., Tildesley, M. J., & Daly, J. M. (2021). Re-parameterization of a mathematical model of African horse sickness virus using data from a systematic literature search. *Transboundary and Emerging Diseases*, 1–11. https://doi.org/10.1111/tbed.14420

CHAPTER 4

Inference for a spatio-temporal model with partial spatial data: African horse sickness virus in Morocco

Note: This chapter is under review as a research article.

Supplementary material for this chapter can be found in Appendix F.

Author contribution: Conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, Writing – original draft, Writing – review & editing.

Inference for a spatio-temporal model with partial spatial data: African horse sickness virus in Morocco

Emma L. Fairbanks[1], Matthew Baylis[2], Janet M. Daly[1], and Michael J. Tildesley[3]

[1]School of Veterinary Medicine and Science, University of Nottingham, Loughborough. LE12 5RD, UK
[2]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Leahurst Campus, Neston, Cheshire, CH64 7TE, UK
[3]The Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research, School of Life Sciences and Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

Abstract African horse sickness virus (AHSV) is a vector-borne virus spread by midges (*Culicoides* spp.). The virus causes African horse sickness (AHS) disease in some species of equid. AHS is endemic in parts of Africa, previously emerged in Europe and in 2020 caused outbreaks for the first time in parts of Eastern Asia. Here we analyse a unique historic dataset from the 1989-1991 emergence of AHS in Morocco in a naïve population of equids. Sequential Monte Carlo and Markov chain Monte Carlo techniques are used to estimate parameters for a spatial-temporal model using a transmission kernel. These parameters allow us to observe how the transmissiblity of AHSV changes according to the distance between premises. We observe how the spatial specificity of the dataset giving the locations of premises on which any infected equids were reported affects parameter estimates. Estimations of transmissiblity were similar at the scales of village (location to the nearest 1.3 km) and region (median area 99 km2), but not province (median area 3000 km^2). This data-driven result could help inform decisions by policy makers on collecting data during future equine disease outbreaks, as well as policies for AHS control.

Keywords: Vector-borne disease, spatio-temporal model, Bayesian inference

1 Introduction

African horse sickness (AHS) is a disease of equids caused by Afican horse sickness virus (AHSV) belonging to the *Reoviridae* family, genus *Orbivirus*. This vector-borne disease, spread between equids by biting midges (*Culicoides* spp.), is endemic in parts of Africa. Historic outbreaks, caused by serotype-9, have occurred in other (non-endemic) regions of Africa, the Middle East and Asia and Europe (Spain) [2, 5, 14, 20]. In July 1987, equids in Spain became infected with AHSV serotype 4, which is believed to have been caused by importation of sub-clinically infected zebras from Namibia. New infections continued to arise until October, when lower temperatures restricted the ability of the virus to spread. Although it was initially hoped that this would eradicate AHSV from the region, the virus successfully overwintered [18, 31]. The next year, the virus persisted in Spain alone, however in 1989 it spread to Portugal and Morocco. The outbreak in Portugal only lasted 13 weeks

due to the introduction of a policy of mass vaccination, strict movement controls and culling of equids housed at the same premises where infected equids were identified. Overall, 206 Portuguese equids were infected and the total cost was estimated at 2 million US dollars [22]. The next year, Spain eradicated the virus after a monovalent vaccine was administered to 38,000 susceptible equids. During the outbreak in Spain, there were 400 infectious cases and a further 900 were culled in an attempt to control the spread of AHSV [19, 25]. The virus was not eradicated from Morocco until 1991 [3]. More recent outbreaks include serotype 2 in Nigeria and Senegal and serotype 7 in Senegal in 2007 [8] and serotypes 2, 4, 6, 8 and 9 in Ethiopia [1] in 2007-2010. Remarkably, in 2020 AHS (serotype 1) occurred for the first time in Thailand and Malaysia [7, 17], this is the furthest east the disease has emerged. This demonstrates the threat this disease continues to pose in both Africa and other parts of the world, including countries historically unaffected by the disease.

In naive populations of horses, mortality due to AHS may exceed 90% in epidemics [7]. There is no specific treatment for animals with AHS but vaccines are available for all nine serotypes. In endemic regions, a live-attenuated vaccine is used. However, there are concerns regarding use of live-attenuated AHSV vaccines because of their capacity for reversion to virulence and transmission by midges, the potential for reassortment between the vaccine and field strains of virus leading to novel viruses and teratogenic effects [21, 33]. Inactivated vaccines avoid these potential drawbacks. Previously, they have been used in Iran in the 1960s and in Spain in 1993 [7] but are no longer commercially available,. The emergence of bluetongue virus (BTV), a closely related virus also transmitted by *Culicoides*, in Europe in 2006 has raised concerns about the potential of AHSV outbreaks in Europe's naïve equine populations [34].

Spatio-temporal models are often used to understand viral outbreaks in livestock species, including BTV [4, 30]. Previously, models have been used to assess the spatio-temporal risk of AHSV, but not transmission between premises [10, 16]. A key difference in our ability to accurately model outbreaks of BTV is the greater quality of data available on distribution and movement of cattle and other livestock species compared with that available for equids in many countries. In this study, we used a spatial dispersal kernel model to analyse data on the emergence of AHSV in Morocco in 1989. Here we know the locations of the premises on which infected equids were identified ('infected premises') but not of the 'uninfected premises'. However, data available on the estimated number of equids in each province is used to approximate their spatial distribution. We estimated parameters given the data at the province, region and village level for the infected premises. This allowed assessment of how the quality of spatial data available influences how well we can model the outbreak.

2 Methods

2.1 Data

Data for the Moroccan outbreak from 1989–1991 described by Baylis et al. [3] were obtained in 1993 by examining case records held at the Ministére de l'Agriculture et de la Réforme

Agraire, Rabat, to obtain the names and locations of all villages in Morocco where cases of AHS were reported. The latitudes and longitudes of these villages were obtained by identifying named villages on 1:50,000 Morocco survey maps.

Various datasets were available as estimates for the population number and distribution of equids. Some gave estimates for individual rural communities based on vaccination records. However, these did not include all communities in the outbreak area. The dataset documents were hard copies translated into French close to the time of the outbreak. This caused issues, with many villages having multiple translations from Arabic to French, which varied across the datasets. The minimum scale at which all equine population estimates were available was the province. Shapefiles for Morocco were available from [11] at four levels, two of which matched the data available. These are referred to here as the provinces (level 2) and regions (level 4). Due to many of the names of regions differing between the historic records and modern shapefiles, the co-ordinates for each infected premises were plotted using QGIS and assigned the region name given in the shapefile these co-ordinates were within. Some cases in Tanger province were found to be in another province called Fahs (not mentioned in the French dataset). The province names were updated to those given in the shapefiles.

The data included an estimate of the number of equids in each province. Due to this not being at the species level, the data of infected animals were simplified so they were not species specific. Therefore, for infected premises, the data available were: date infected, province (all premises), region (270/271 premises), latitude and longitude of village (243/271 premises), number of equids infected and total number of equids on premises. No information was available for the uninfected premises.

2.2 Spatial distribution

Uninfected premises are assumed to have the same distribution of the total number of equids on each premises as the infected premises. Therefore, a gamma distribution is fitted to the total number of equids on infected premises, giving a shape parameter of 1.85 and scale parameter of 1.24. The total number of equids in each province was divided into premises such that the total number of equids on uninfected premises followed the same gamma distribution as the infected premises. As we only know the province of these simulated uninfected premises they are randomly placed in the shapefile of the province they are in. Here an additional shapefile was made, using QGIS, combining Tanger and Fahs to match the historic data. The infected farms are placed randomly in the shapefile of the province or region they are in for the province or region level of spatial specificity, respectively. As the latitude and longitude co-ordinates for the village of infected premises are given to the nearest minute, we randomly place these within a circle of radius 30 seconds from the co-ordinates of the village. This also avoids multiple premises being assigned the same location.

2.3 Transmission model

The incubation period for AHSV determined from the literature (Fairbanks *et al.*, submitted), varied from 2 days to 11 days with mode 3 days. The latent period varied from 3

days to 10 days with mode 4. Delays between the onset of clinical signs, their observation, a veterinarian officially diagnosing and culling taking place has potential to vary depending on location, day and management of premises. Due to the variability in these factors, two premises infected on the same day could potentially be reported several days apart. Therefore we do not consider the latent period between premises in the model.

A kernel is used to describe the spread of AHSV between premises. For each spatial distribution the distance between premises is calculated using the haversine function. We only calculate the distance between infected premises and all other premises to reduce computational costs.

The force of infection (λ) on farm i from farms j on day t is calculated as

$$\lambda_i(t) = \beta N^i \sum_{j \neq i} K(d_{ij}) N^j I_j(t), \tag{1}$$

where N^i is the number of equids on premises i, d_{ij} is the distance between premises i and j, K is the kernel-function, and $I_j(t)$ is 1 if premises j is infected on day t and 0 otherwise. β is the transmission parameter. The likelihood is therefore given as

$$L = \prod_{j \in U} \left(\prod_{t} e^{-\lambda_j(t)} \right) \prod_{j \in I} \left(\left(\prod_{t=t_0}^{t_{inf}-1} e^{-\lambda_j(t)} \right) \left(1 - e^{-\lambda_j(t_{inf})} \right) \right), \tag{2}$$

where U is the set of premises with no report of infected equids during the outbreak, I is the set of premises with reported infected equids during the outbreak, t_0 is the start of the outbreak and t_{inf} is the time at which infection is detected on an infected premises. This method does not require the model to be simulated, therefore reducing computational costs.

To identify the most appropriate kernel model, PubMed was searched using the search terms (disease OR virus) AND ∼model AND kernel AND ∼transmission'. Eligible studies included fitting parameters for a distance-based kernel model of disease spread between livestock premises. Supplementary file 1 describes results from the literature review. The selected kernel is given as

$$K = exp(-\alpha d^{\epsilon}). \tag{3}$$

2.4 Parameter estimation

Szmaragd et al. [30] attempted to fit this kernel (Equation 3) to the spread of BTV in Northern Europe during 2006. Here multiple values of ϵ were fit to the data with $\epsilon = 2$ yielding the best fit, therefore we set ϵ to 2. The sequential Monte Carlo-Markov chain Monte Carlo (SMC-MCMC) approach was applied as described in Supplementary file 2. As the likelihood was not numerically stable (converging to 0), the log-likelihood was given as

$$ln(L) = -\sum_{j \in U} \left(\sum_{t} \lambda_j(t) \right) - \sum_{j \in I} \left(\left(\sum_{t=t_0}^{t_{inf}-1} \lambda_j(t) \right) - ln\left(1 - e^{-\lambda_j(t_{inf})} \right) \right), \tag{4}$$

was used.

Initially, to derive our prior, we consider the parameter values fitted by Szmaragd et al. [30] to explain the spread of BTV-8 in Northern Europe during 2006 for the kernel (Equation 3. Our kernel is less informed than Szmaragd et al. [30] as it does not include parameters such as the probability of acquiring infection, which evidence suggests may be different for AHSV and BTV [6].
Our parameter priors are therefore calculated based on the BTV parameters as:

$$\beta = \text{the probability of acquiring infection x } \frac{\hat{\alpha}}{\sqrt{\pi}},$$
$$\alpha = \hat{\alpha}^2,$$

where $\hat{\alpha}$ is thew kernal parameter (α) from Szmaragd et al. [30]. This gives values of $\beta = 0.0108$ and $\alpha = 0.0012$. To initialise the model, Latin hypercube sampling (LHS) is used to select parameters from a uniform distribution with minimum of 0.5x and maximum of 2x the calculated values for β and α. This yields a lower bound of 0.0054 and 0.0006 and upper bound of 0.0216 and 0.0024 for β and α, respectively.

As the spatial distribution of uninfected premises is estimated, it may affect the parameter estimation. Therefore, each iteration we generate multiple spatial distributions that each have a different set of parameters depending on the results of the previous iterations. We use a modified version of Ripley's L function (Supplementary file 3) to determine how the spatial homogeneity of the distribution of randomly distributed uninfected premises varied around the infected premises across spatial distributions. If the spatial distributions are similar, this will cause less variation during parameter estimation, therefore fewer iterations will be needed than if they are very different. As little variability was observed between spatial distributions, we conclude that a large number of different spatial distributions is not necessary at each MCMC iteration (Supplementary file 3).

3 Results

It was found that disease outbreaks were widespread across the country, with the greatest number of reported cases in the northwest (1989–1990). Location data for 1991 were sparse, but it is possible to observe the increased range of the virus. In 1991, there were significant numbers of infected equids in Marrakesh province. Despite having a large equine population, there were no cases reported west of the Atlas Mountains. During the first year of the outbreak, 271 infected premises were recorded, resulting in a total of 518 cases and an additional 106 equids culled. Cases were reported in the provinces of Larache, Tanger and Tetouan.

Initially during the parameter estimation using the latitude and longitude of villages, we observe a large deviation from the prior (Supplementary figure S3). The first 100 iterations are removed as burn-in. After this, parameters appear to follow normal distributions (Table 1). As α appears somewhat skewed, a log-normal distribution was fitted, but this yielded

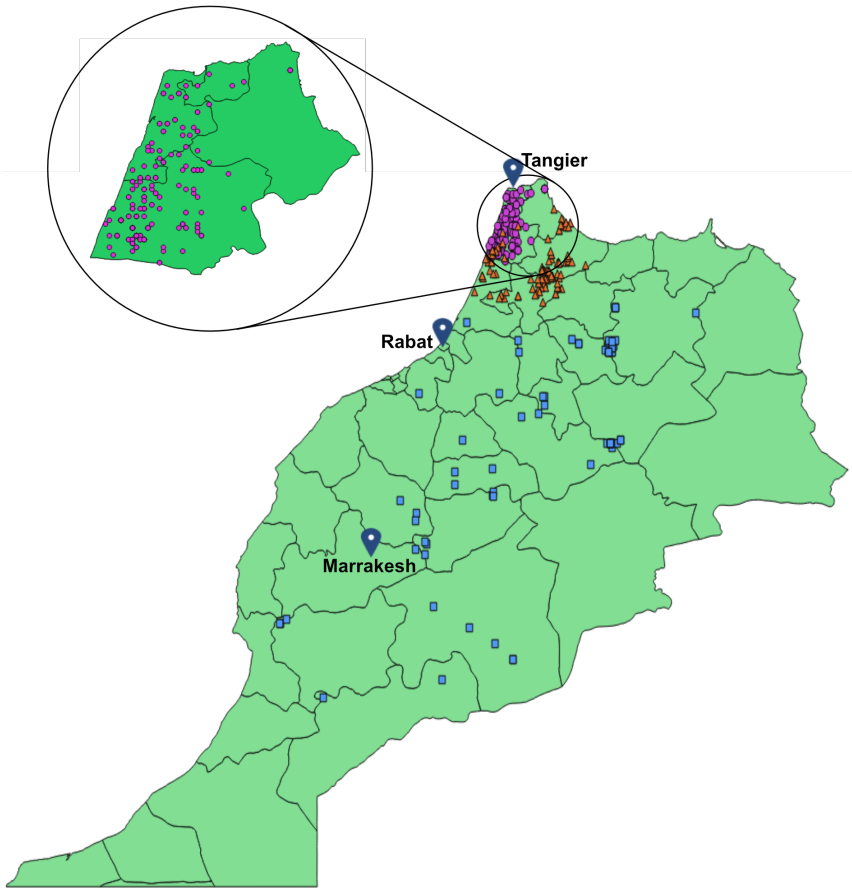

Figure 1: Location of AHS cases in Morocco for 1989 (pink dots), 1990 (orange triangles) and available 1991 (blue squares). QGIS was used to make this figure [23].

the same result.

As the spatial specificity decreases, the variance of accepted parameter values increases (Supplementary figures 3-5). The region data provide parameter estimates much more similar to the village data than the province data (Table 1).

Another more simple approach to analysing the patterns of transmission involves calculating the distances between the premises where latitude and longitude of the villages of infected premises are given and calculating the minimum distance of each premises from a previously infected premises (Figure 2a). We observe that most of these distances are short ($<$5 km), although there were some cases of transmission occurring over larger distances. Comparing outputs from this and our transmission kernel shows that our kernel also predicts most of the transmission occurring locally (Figure 2b).

The prior and posterior values for the kernel shape parameter were similar at the province level (Figure 2b). However, the posterior values for AHSV at the village and region levels suggest that AHSV transmission during the Moroccan outbreak occurred over shorter dis-

Spatial specificity	Parameter	Mean	SD
Village	β	$1.00\text{x}10^{-6}$	$5.64\text{x}10^{-8}$
	α	$2.86\text{x}10^{-3}$	$9.42\text{x}10^{-5}$
Region	β	$1.23\text{x}10^{-6}$	$2.18\text{x}10^{-7}$
	α	$3.45\ \text{x}10^{-3}$	$7.09\text{x}10^{-4}$
Province	β	$4.01\text{x}10^{-7}$	$1.07\text{x}10^{-7}$
	α	$1.09\text{x}10^{-3}$	$3.31\text{x}10^{-4}$

Table 1: Posterior values of parameter estimates for each spatial specificity. A normal distribution was fitted to accepted parameter values after burn-in (first 100 iterations). The mean and standard deviation (SD) of the normal distributions are given.

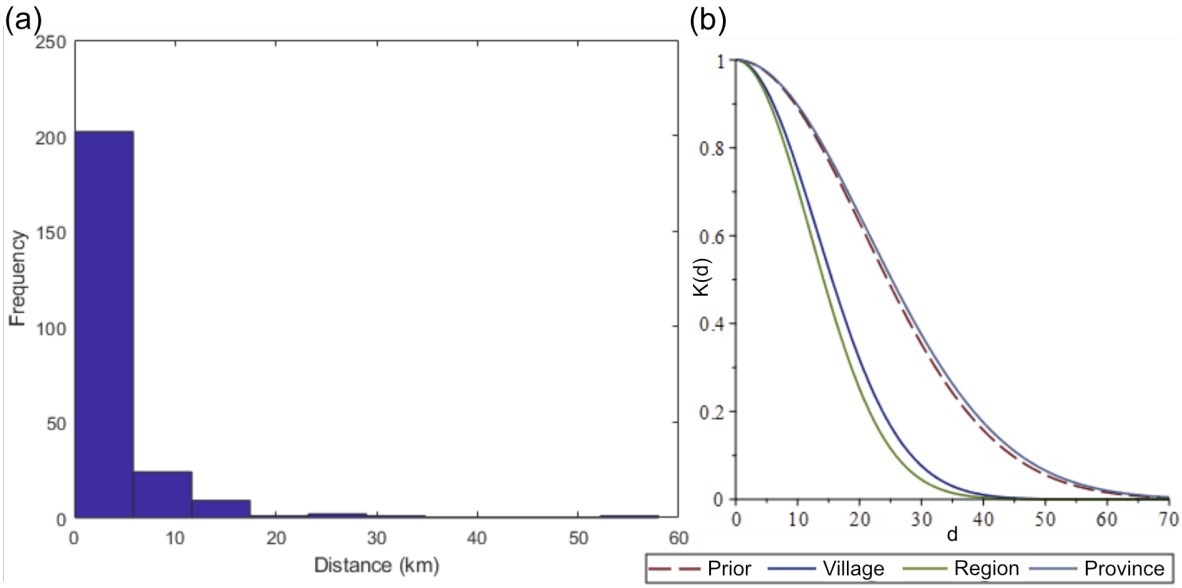

Figure 2: (a) Histogram of the minimum distance between each premises and a previously infected premises. (b) The (bluetongue virus) prior (red dashed line) and (African horse sickness virus) posterior for the village (blue line), region (green line) and province (grey line) kernel shape parameter.

tances than during the 2001 BTV outbreak in northern Europe, from which the prior values were derived.

4 Discussion

In this paper we were able to parameterise a model for the 1989 emergence of AHSV in Morocco for three different spatial specificities of data on infected equids (village, region and province). Data-informed models are important for stakeholders with responsibility for controlling outbreaks in naïve populations, which are increasingly likely with climate change and globalisation [24]. Previously, a lack of data has prohibited modelling the transmission of AHS between premises, but this unique dataset allowed a first attempt at this. One of the biggest challenges encountered was the computational demand of this large individual-scale model. Missing data were compensated for using multiple random distributions. Here the Ripley's L function was used to determine the spatial variation between these random distributions. Observing that there was little variance in spatial homogeneity between premises near infected premises allowed the number of SMC steps to be reduced each iteration. Previous research has shown the optimal acceptance rate is 0.234 under quite general conditions [12]. Further research has indicated that this acceptance rate does seem to perform well for approximate finite-dimensional situations where the number of parameters jointly estimated is as few as five. This does not apply when estimating fewer parameters. For example, it has been found that for one normally distributed parameter the optimal acceptance rate is approximately 0.44 [26]. Here, we accepted up to 30% of the parameter sets each iteration, dependent upon their log-likelihood.

The number of equids on 'uninfected premises' was assumed to follow the same distribution of those on premises with a reported infection. Gubbins et al. [13] suggested that the principle mechanism behind AHSV transmission is vector dispersal [13]. Analysis of the 1989 Morocco outbreak showed most of the transmission occurred over short distances, consistent with studies of *Culicoides* dispersal [15, 27]. However, not all transmission may be due to the flight of midges. Other factors, for example new importations of infectious vectors or the movement of animals or reservoir hosts could be responsible. The parameterised kernel shape parameter (α) suggested AHSV transmission during this outbreak occurred over shorter distances than the 2001 BTV outbreak in northern Europe (used as the prior). This does not necessarily imply that AHSV lacks BTV's potential to spread over larger distances. Rather, this could be caused by equine premises being more clustered compared to cattle and sheep holdings, and therefore most of the transmission is local.

Tildesley et al. [32] showed that spatial structure can be important for predictions of emergence and population-scale dynamics, however this structure is mostly subsumed in the parameterisation when considering the effectiveness of control strategies. Here we assume that the premises stay infected after infected equids are culled as local infected midges will remain present for a period. However, there is uncertainty regarding midge behaviour at the time after culling. Elbers and Meiswinkel [9] quantified host preferences of *Culicoides* during early summer in the Netherlands. Generally, there was a strong correlation between attack rate and host size with a cow attracting the greatest proportion (62.4%) of the *Culicoides* captured, followed by a Shetland pony (29.2%) and then a sheep (8.4%). However, similar numbers of *C. dewulfi* were collected from the cow and the pony and *C. obsoletus* numbers where evenly distributed among the three host species. In contrast, Schmidtmann et al. [29]

captured *C. obsoletus* in comparable numbers from calves and sheep but significantly smaller numbers were found on ponies [29]. Therefore, when equids are removed from a premises by culling, how far the *Culicoides* may travel to seek alternative hosts may depend on the species of *Culicoides* and the distribution and density of other species locally.

One of the most important features of vector-borne diseases not addressed in this model is climate impacts as climate data from the time of the AHS outbreak were not available for the region of this study. Climate factors will impact both the abundance of *Culicoides* and the dynamics of virus transmission [6, 28]. Although climate data are not available to inform this model, the within-premises dynamics may help provide information on vector capacity throughout the outbreak. For example, if more secondary infections are reported (before culling) for a period within the outbreak, we could assume that the vector capacity was higher at the time.

In conclusion, although vector capacity is likely to vary between geographical regions this model may help inform the dynamics of potential outbreaks upon emergence. Overall, we found that the village and region datasets generated similar parameter estimates, whereas the province level data did not. Data on the location of equids are often sparse; for example, reports of equine influenza in 2019 were given for each nomenclature of territorial unit 2 (NUTS2) of the UK. The median area of a NUTS2 region in the UK is 4,788 km^2, which is most comparable to the median province area for the three provinces of Morocco with infected equids in 1989 (2,653 km^2). This suggests that, in the unfortunate circumstance of an AHSV outbreak in the UK, for example, spatial data would need to be collected at a finer scale to predict the population-scale spatial dynamics of the virus from a knowledge of the individual-level dynamics.

Acknowledgements ELF acknowledges support from the Nottingham BBSRC Doctoral Training Partnership. MB acknowledges colleagues at the Ministere d'Elevage in Rabat in 1993 for providing access to their records of the AHS outbreak in Morocco in 1989-1991.

Ethics statement The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required.

Conflict of interest statement The authors declare that they have no conflict of interest.

Data availability An example randomly generated region-level spatial distribution and code are available on the University of Nottingham data sharing platform (https://rdmc.nottingham.ac.uk/).

Author contributions
Emma L. Fairbanks: Conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, Writing – original draft, Writing – review & editing.

Matthew Baylis: Conceptualization, data curation, methodology, resources, Writing – review & editing.
Janet M. Daly: Conceptualization, methodology, project administration, supervision, Writing – original draft, Writing – review & editing.
Michael J. Tildesley: Conceptualization, methodology, software, supervision,Writing – review & editing.

ORCID ID
Emma L. Fairbanks 0000-0002-1598-962X
Michael J. Tildesley 0000-0002-6875-7232
Matthew Baylis 0000-0003-0335-187X
Janet M. Daly 0000-0002-1912-4500

References

[1] Aklilu, N., Batten, C., Gelaye, E., Jenberie, S., Ayelet, G., Wilson, A., Belay, A., Asfaw, Y., Oura, C., Maan, S., Bachanek-Bankowska, K. and Mertens, P. P. C. [2014], *Transbound Emerg Dis* **61**(2), 185–192.

[2] Anwar, M. and Qureshi, M. [1973], 'Control and eradication of African horse sickness in Pakistan', *Dev. Biol.* **199**, 225–228.

[3] Baylis, M., Hasnaoui, H. E. L., Bouayoune, H., Touti, J. and Mellor, P. S. [1997], 'The spatial and seasonal distribution of African horse sickness and its potential Culicoides vectors in Morocco', *Med Vet Entomol* **11**(3), 203–212.

[4] Boender, G. J., Hagenaars, T. J., Elbers, A. R., Gethmann, J. M., Meroc, E., G, H. and De Koeijer, A. A. [2014], 'Confirmation of spatial patterns and temperature effects in bluetongue virus serotype-8 transmission in NW-Europe from the 2007 reported case data', *Vet Res* **45**(1), 75.

[5] Carpenter, S., Mellor, P. S., Fall, A. G., Garros, C. and Venter, G. J. [2017], 'African horse sickness virus: history, transmission, and current status', *Annu Rev Entomol* **62**, 343–358.

[6] Carpenter, S., Wilson, A., Barber, J., Veronesi, E., Mellor, P., Venter, G. and Gubbins, S. [2011], 'Temperature dependence of the extrinsic incubation period of orbiviruses in Culicoides biting midges', *PloS one* **6**(11), e27987.

[7] Castillo-Olivares, J. [2021], 'African horse sickness in Thailand: Challenges of controlling an outbreak by vaccination', *Equine Vet J* **51**(1), 9–14.

[8] Diouf, N. D., Etter, E., Lo, M. M., Lo, M. and Akakpo, A. [2013], 'Outbreaks of African horse sickness in Senegal, and methods of control of the 2007 epidemic', *Vet Rec* **172**(6), 152–152.

[9] Elbers, A. and Meiswinkel, R. [2015], 'Culicoides (Diptera: Ceratopogonidae) and livestock in the Netherlands: comparing host preference and attack rates on a Shetland pony, a dairy cow, and a sheep', *J Vector Ecol* **40**(2), 308–317.

[10] Faverjon, C., Leblond, A., Hendrikx, P., Balenghien, T., de Vos, C., Fischer, E. and De Koeijer, A. A. [2015], 'A spatiotemporal model to assess the introduction risk of African horse sickness by import of animals and vectors in France', *BMC Vet Res* **11**(1), 1–15.

[11] *GADM* [2018], https://gadm.org/download_country_v3.html. Accessed: 2019-12-02.

[12] Gelman, A., Gilks, W. R. and Roberts, G. O. [1997], 'Weak convergence and optimal scaling of random walk Metropolis algorithms', *Ann Appl Probab* **7**(1), 110–120.

[13] Gubbins, S., Richardson, J., Baylis, M., Wilson, A. and Abrahantes, J. [2014], 'Modelling the continental-scale spread of Schmallenberg virus in Europe:Approaches and challenges', *Prev Vet Med* **116**(4), 404–411.

[14] Howell, P. [1960], 'The 1960 epizootic of African horse sickness in the Middle East and SW Asia', *Journal S Afr Vet Assoc* **31**(3), 329–334.

[15] Kluiters, G., Swales, H. and Baylis, M. [2015], 'Local dispersal of palaearctic culicoides biting midges estimated by mark-release-recapture', *Parasite Vector* **8**(1), 1–9.

[16] Lo Iacono, G., Robin, C. A., Newton, J. R., Gubbins, S. and Wood, J. L. [2013], 'Where are the horses? with the sheep or cows? uncertain host location, vector-feeding preferences and the risk of African horse sickness transmission in Great Britain', *J Roy Soc Interface* **10**(83), 20130194.

[17] Lu, G., Pan, J., Ou, J., Shao, R., Hu, X., Wang, C. and Li, S. [2020], 'African horse sickness: Its emergence in Thailand and potential threat to other Asian countries', *Transbound Emerg Dis* **67**(5), 1751–1753.

[18] Lubroth, J. [1988], 'African horsesickness and the epizootic in Spain 1987', *AAEP PROCEEDINGS* **33**, 871–877.

[19] Mellor, P. [1993], 'African horse sickness: transmission and epidemiology', *Vet Res* **24**(2), 199–212.

[20] Mellor, P. S. and Boorman, J. [1995], 'The transmission and geographical spread of African horse sickness and bluetongue viruses', *Ann Trop Med Parasitol* **89**(1), 1–15.

[21] Mellor, P. S. and Hamblin, C. [2004], 'African horse sickness', *Vet Res* **35**(4), 445–466.

[22] Portas, M., Boinas, F. S., Oliveira, J., Sousa, E. and Rawlings, P. [1999], 'African horse sickness in Portugal: a successful eradication programme', *Epidemiol Infect* **123**(2), 337–346.

[23] *QGIS Geographic Information System. QGIS Association* [2021], http://www.qgis.org.

[24] Rocklöv, J. and Dubrow, R. [2020], 'Climate change: an enduring challenge for vector-borne disease prevention and control', *Nat Immunol* **21**(5), 479–483.

[25] Rodriguez, M., Hooghuis, H. and Castano, M. A. [1992], 'African horse sickness in Spain', *Vet Micro* **33**(1-4), 129–142.

[26] Rosenthal, J. S. [2011], 'Optimal proposal distributions and adaptive MCMC', *Handbook of Markov Chain Monte Carlo* **4**(10.1201).

[27] Sanders, C. J., Harrup, L. E., Tugwell, L. A., Brugman, V. A., England, M. and Carpenter, S. [2017], 'Quantification of within-and between-farm dispersal of Culicoides biting midges using an immunomarking technique', *J Appl Ecol* **54**(5), 1429–1439.

[28] Sanders, C. J., Shortall, C. R., Gubbins, S., Burgin, L., Gloster, J., Harrington, R., Reynolds, D. R., Mellor, P. S. and Carpenter, S. [2011], 'Influence of season and meteorological parameters on flight activity of Culicoides biting midges', *J Appl Ecol* **48**(6), 1355–1364.

[29] Schmidtmann, E. T., Jones, C. J. and Gollands, B. [1980], 'Comparative host-seeking activity of culicdides (Diptera: Ceratopogonidae) attracted to pastured livestock in central New York state, USA', *J Med Entomol* **17**(3), 221–231.

[30] Szmaragd, C., Wilson, A. J., Carpenter, S., Wood, J. L. N., Mellor, P. S. and Gubbins, S. [2009], 'A modeling framework to describe the transmission of bluetongue virus within and between farms in Great Britain', *PloS one* **4**(11), e7741.

[31] Thompson, G. M., Jess, S. and Murchie, A. K. [2012], 'A review of African horse sickness and its implications for Ireland', *Irish Vet J* **65**(1), 9.

[32] Tildesley, M. J., House, T. A., Bruhn, M. C., Curry, R. J., O'Neil, M., Allpress, J. L. E., Smith, G. and Keeling, M. J. [2010], 'Impact of spatial clustering on disease transmission and optimal control', *P Natl Acad Sci USA* **107**(3), 1041–1046.

[33] Van Rijn, P. A., Maris-Veldhuis, M. A., Grobler, M., Wright, I. M., Erasmus, B. J., Maartens, L. H. and Potgieter, C. A. [2020], 'Safety and efficacy of inactivated African horse sickness (ahs) vaccine formulated with different adjuvants', *Vaccine* **38**(45), 7108–7117.

[34] Zientara, S., Weyer, C. T. and Lecollinet, S. [2015], 'African horse sickness', *University of Pretoria* . Available at:https://repository.up.ac.za/handle/2263/51662.

CHAPTER 5

SARS-CoV-2 infection in UK university students: lessons from September-December 2020 and modelling insights for future student return

Note: This chapter is a published article:

Enright J, Hill EM, Stage HB, Bolton KJ, Nixon EJ, Fairbanks EL, Tang ML, Brooks-Pollock E, Dyson L, Budd CJ, Hoyle RB, Schewe L, Gog, JR and Tildesley, MJ. SARS-CoV-2 infection in UK university students: lessons from September–December 2020 and modelling insights for future student return. Royal Society open science. 2021;8(8):210310.

Royal society publishing grants authors full rights to reuse their articles as part of a thesis.

This work presented as part of a SAGE teach-in, hosted and chaired by GO-Science, as well as being presented at SAGE and utilised to directly inform government decision making when the government road map to lift lockdown was announced on 22nd February 2021. The research was released as one of the SAGE papers on that date.

Author contribution: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualisation.

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: 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.* **8**: 210310. https://doi.org/10.1098/rsos.210310

Received: 23 February 2021
Accepted: 16 July 2021

Subject Category:
Science, society and policy

Subject Areas:
mathematical modelling/health and disease and epidemiology

Keywords:
epidemic modelling, pandemic modelling, COVID-19, SARS-CoV-2, higher education

Authors for correspondence:
Jessica Enright
e-mail: jessica.enright@glasgow.ac.uk
Edward M. Hill
e-mail: Edward.Hill@warwick.ac.uk
Julia R. Gog
e-mail: jrg20@cam.ac.uk
Michael J. Tildesley
e-mail: M.J.Tildesley@warwick.ac.uk

†These authors contributed equally to this work.

THE ROYAL SOCIETY PUBLISHING

SARS-CoV-2 infection in UK university students: lessons from September–December 2020 and modelling insights for future student return

Jessica Enright[1,†], Edward M. Hill[2,3,†], Helena B. Stage[3,4], Kirsty J. Bolton[5], Emily J. Nixon[3,6], Emma L. Fairbanks[5,7], Maria L. Tang[3,8], Ellen Brooks-Pollock[3,9], Louise Dyson[2,3], Chris J. Budd[10], Rebecca B. Hoyle[11], Lars Schewe[12], Julia R. Gog[3,8] and Michael J. Tildesley[2,3]

[1]School of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK
[2]The Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, School of Life Sciences and Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK
[3]Joint UNIversities Pandemic and Epidemiological Research, https://maths.org/juniper/
[4]Department of Mathematics, The University of Manchester, Oxford Road, Manchester, UK
[5]Centre for Mathematical Medicine and Biology, School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, UK
[6]Veterinary Public Health, Bristol Veterinary School, University of Bristol, Bristol, UK
[7]School of Veterinary Medicine and Science, University of Nottingham, Loughborough, UK
[8]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK
[9]Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, UK
[10]School of Mathematical Sciences, University of Bath, Claverton Down, Bath, UK
[11]School of Mathematical Sciences, University of Southampton, Southampton, UK
[12]University of Edinburgh, School of Mathematics, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh, UK

JE, 0000-0002-0266-3292; EMH, 0000-0002-2992-2004; HBS, 0000-0001-9938-8452; KJB, 0000-0003-0487-4701; EJN, 0000-0002-1626-9296; ELF, 0000-0002-1598-962X; MLT, 0000-0002-9671-8302; EB-P, 0000-0002-5984-4932; LD, 0000-0001-9788-4858; CJB, 0000-0003-4536-1662; RBH, 0000-0002-1645-1071; LS, 0000-0002-3778-262X; JRG, 0000-0003-1240-7214; MJT, 0000-0002-6875-7232

In this paper, we present work on SARS-CoV-2 transmission in UK higher education settings using multiple approaches to assess the extent of university outbreaks, how much those outbreaks may have led to spillover in the community, and the expected effects of control measures. Firstly, we found that the distribution of outbreaks in universities in late 2020 was consistent with the expected importation of infection

from arriving students. Considering outbreaks at one university, larger halls of residence posed higher risks for transmission. The dynamics of transmission from university outbreaks to wider communities is complex, and while sometimes spillover does occur, occasionally even large outbreaks do not give any detectable signal of spillover to the local population. Secondly, we explored proposed control measures for reopening and keeping open universities. We found the proposal of staggering the return of students to university residence is of limited value in terms of reducing transmission. We show that student adherence to testing and self-isolation is likely to be much more important for reducing transmission during term time. Finally, we explored strategies for testing students in the context of a more transmissible variant and found that frequent testing would be necessary to prevent a major outbreak.

1. Introduction

The global spread of SARS-CoV-2 has resulted in widespread usage of social distancing measures and non-pharmaceutical interventions (NPIs) to inhibit the spread of infection. Enactment of nationwide lockdowns has resulted in the closure of workplaces, pubs and restaurants, restricted leisure activities and impacted the education sector.

Measures brought in when entering the first nationwide lockdown in the UK in March 2020 included closure of higher education establishments, such as universities, to most in-person activities. Face-to-face teaching was mostly suspended, with delivery of the remainder of the 2019/2020 academic year taking place via online delivery.

Higher education in the UK comprises a large population of students, with over 2.3 million higher education students enrolled in the 2018/2019 academic year across over 160 higher education providers [1] (universities, essentially). This results in a sizeable movement of students nationwide at the beginning and end of academic terms (in addition to international student travel). In the context of an ongoing disease outbreak, the migration of students can contribute to increased population mobility, with an associated need for careful management in order to minimize the risk of seeding outbreaks both in universities and in the wider community.

Ahead of the 2020/2021 academic year, there was significant uncertainty around whether students would be able to return to face-to-face teaching and what policies would be put in place in order to mitigate risk. This prompted action to build a foundation of knowledge such that appropriate policies could be put in place to facilitate students returning safely to universities. From 15 to 17 June 2020, a Virtual Study Group on 'Unlocking Higher Education Spaces' was hosted by the Virtual Forum for Knowledge Exchange in the Mathematical Sciences (V-KEMS), looking at how mathematical approaches could inform the reopening of higher education spaces to students while minimizing risk. A working paper was subsequently released in July 2020 [2].

Building on the discussion that took place at the June 2020 Study Group, two virtual events (taking place on 28 July 2020 and 4 August 2020, respectively) investigated the application of mathematical tools and models to various issues linked to the challenges of reopening higher education. These events were run as part of the Isaac Newton Institute Infectious Dynamics of Pandemics Research Programme [3]. After these events, a working group continued to meet virtually on a weekly basis, consisting of participants from several institutions.

Mathematical modelling approaches informed by data, have been a valuable tool used to inform policy decisions linked to the subsequent operation of higher education in the midst of a pandemic. In order to guide these decisions, in this paper, we have investigated contributing factors to within-institution spread and how transmission interplays with the wider community. This study starts with a set of observational analyses based on data from the first term of the 2020/2021 academic year. This is followed by prospective modelling of control measures that were under consideration for the full return of UK higher education students in January 2021.

The work presented in this paper is the outcome of bringing together the expertise from these multiple research groups, and pooling our analyses using both statistical and modelling methods. Several conclusions emerge from this work both in understanding the observations from Autumn 2020, and also making recommendations for future actions:

(1) The overall distribution of outbreaks in universities in autumn term 2020 were consistent with expected importations from taking a student intake from the wider community, so that universities reflect the community disease prevalence at the start of term.

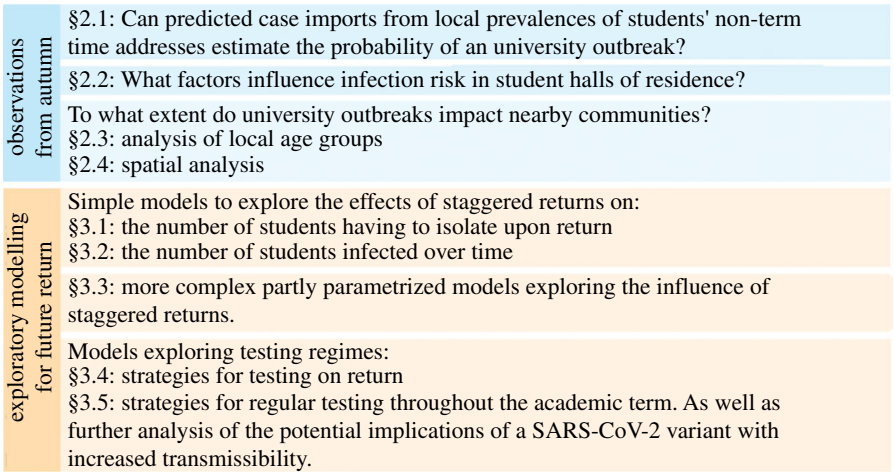

Figure 1. Overview of the structure of the article.

(2) Larger halls of residence pose higher risks for larger attack rates, and segmentation into smaller households within halls is unlikely to be able to mitigate this.
(3) The picture of transmission from universities to their local communities is complex. While spillover inevitably can occur, sometimes even large outbreaks in universities do not give any corresponding signal in their wider neighbouring communities.
(4) The proposed strategy of staggering future returns appears to be of somewhat mixed and limited value. While it could reduce the need for self-isolation on return under low prevalence, these benefits could be diminished or even reversed in the context of high background prevalence.
(5) While a staggered return could reduce the peak of any outbreak during term, staggering on its own will not substantially reduce the total attack rate over a whole term: staggering may act mainly to delay the outbreak to later in the term.
(6) The level of student adherence to testing and isolation is likely to have a far larger effect than any subtleties between different staggered return regimes.
(7) While it is likely that asymptomatic testing programmes did help to prevent large outbreaks in university settings in autumn 2020, extremely frequent testing (every 3 days) would be needed to prevent a major outbreak under plausible parameters for the B.1.1.7 variant (WHO variant label 'Alpha').

The structure of the remainder of this paper is as follows. In §2, we summarize the understanding and learning from the observed patterns of SARS-CoV-2 from autumn term 2020, looking at the dynamics of wider community transmission including the importation of cases to universities at the start of term, and the spillover of transmissions from universities to the wider community during the course of the term. This section also looks at the dependence on the infection dynamics within universities of the structures of halls of residence and student households. In §3, we look at several exploratory models for the future return of students, in particular looking at the impact of different strategies for staggering this return, and of asymptomatic testing on return. In §4, we draw some further conclusions from this work and make some policy recommendations. For a more detailed overview, see figure 1.

2. Observations from autumn term 2020

Higher education institutions in the UK largely reopened to students for the 2020/2021 academic year. This led to an influx of students from across the UK and world, brought together in residential, academic and social settings. In the first term, under the government advice at the time [4], most higher education establishments offered blended online and face-to-face learning. Prior to the beginning of the academic year, students resident in housing of multiple occupancy—and in particular students in residential halls—were identified as being at high risk of transmitting SARS-CoV-2 infection [5,6].

The return of students to universities in the autumn term occurred at a time when SARS-CoV-2 cases were growing in the UK. Local lockdowns came into force in areas with greatest risk leading to an increase in restrictions on travel, business openings and between-household socializing. In addition,

countrywide lockdowns were imposed in Wales from 23 October 2020 to 9 November 2020 and in England from 5 November 2020 to 2 December 2020. Many universities offered testing regimes in an attempt to further control outbreaks. In an attempt to segment interactions and reduce transmission risk within halls, many universities assigned students in residential halls to households based on the use of shared facilities such as kitchens and bathrooms under government guidance [4]. These households were intended to function similarly to households in the community; with many restrictions on socializing beyond these household members, and requirements for the entire household to isolate for up to 14 days if a member displayed symptoms of COVID-19 or received a positive SARS-CoV-2 test. Despite the control measures taken, outbreaks of varying sizes were seen in many UK higher education institutions in the first term, prompting concern about the possibility of spillover into the community.

In this section, we use data from the first term of the 2020/2021 academic year to investigate the factors that may have contributed to the observed outbreaks within higher education institutions and to examine any evidence of further transmission between higher education institutions and the wider community. Firstly, we consider the mass migration of students from across the UK at the beginning of term and how well this may explain the occurrence of outbreaks seen across universities (§2.1). We then use data available from a particular university and investigate the role of accommodation structure upon transmission, by considering the relationship of residential hall sizes and household sizes within halls to attack rates (§2.2). To investigate spillover from higher education to the community, we investigate case data by age (henceforth 'age-stratified') from areas very close to English universities to determine whether there is any evidence of spillover from student age groups to other age groups (§2.3). We also consider total case data stratified across a wider spatial scale to search for signs of spillover from areas with a high concentration of student residents to geographically nearby areas without high concentrations of students (§2.4).

2.1. Start of term: transmission from the community

Although many universities experienced outbreaks at the beginning of the 2020/2021 academic year, there was significant variation in the number of confirmed SARS-CoV-2 cases between institutions. We explore the extent to which the estimated incoming numbers of infected students could explain the observed distribution of outbreaks in the early weeks of the autumn term across UK universities.

2.1.1. Data and methods

To estimate the number of incoming infected students for each university at the beginning of the 2020/2021 academic year, we combined Office for National Statistics (ONS) infection survey data on the proportion of the community testing positive (prevalence) via polymerase chain reaction (PCR) to SARS-CoV-2 by region with data from the higher education Statistics Agency (HESA) on home and term-time postcodes for the 2018/2019 cohort of students [7]. The prevalence (via PCR) on 25 September 2020 was used to estimate the number of students from each home postcode that were infected at the start of term (rounded to the nearest integer). It was assumed that international students from countries with high case numbers would be placed in effective quarantine and were thus discounted for the purpose of this analysis. Outbreak data were drawn from the University and College Union (UCU) dashboard in November 2020 [8]. After omitting data with obvious quality issues, data for 72 universities were available. We defined a large outbreak as 200 or more cumulative cases reported on the UCU dashboard by 18 or 19 November 2020 (these case numbers obtained relate to various dates in November since updates were not daily or uniform).

To estimate the probability of a large outbreak, universities were binned by the estimated number of PCR-positive students in bin widths of 10, and the fraction of universities in each bin that experienced an outbreak was calculated based on the observed data.

We also considered a simple probabilistic model for the outbreak probability $\mathcal{P}$ based only on incoming PCR-positive students, $\mathcal{P} = 1 - p^n$, where n corresponds to the initial number of PCR-positive students, and the extinction probability, p, is the probability that an incoming infection fails to seed an outbreak. The probabilities of each incoming infection seeding an outbreak are assumed to be independent of each other. The extinction probability, p, was inferred via maximum likelihood from the observed outbreak data (see appendix A).

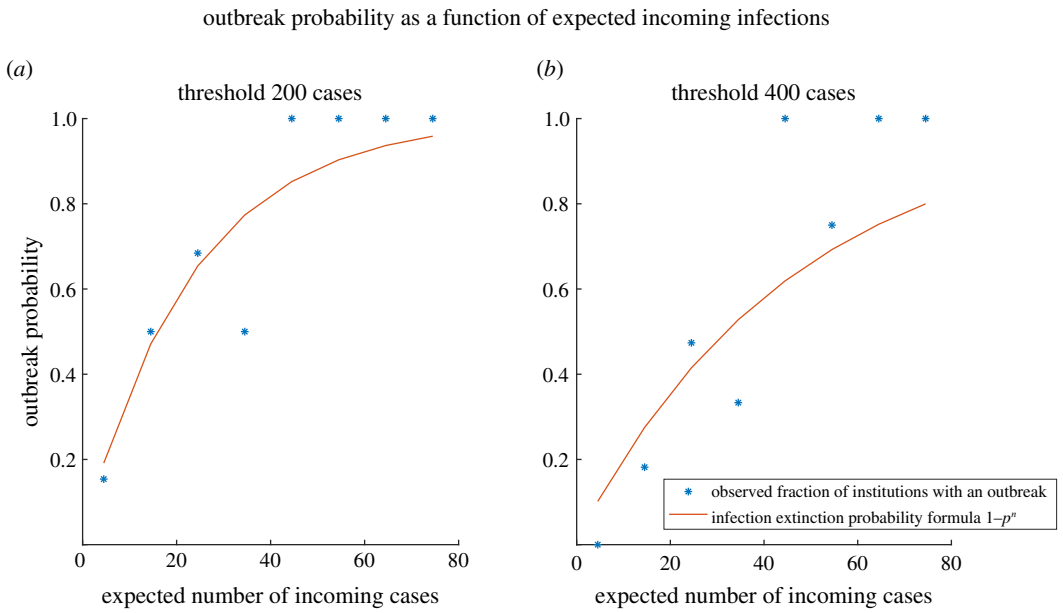

Figure 2. Observed fraction of institutions having an outbreak (*), binned by expected number of incoming cases, and theoretical outbreak probability $\mathcal{P}$ (solid line): for a threshold of 200 cases (*a*) and 400 cases (*b*).

2.1.2. Results

The observed fraction of universities experiencing an outbreak appeared to be broadly consistent with the simple probabilistic model (figure 2), with a fitted extinction probability of $p = 0.958$ (95% confidence interval [0.945, 0.972]). Repeating the analysis and fitting the simple model using a more stringent threshold of 400 cases returned an extinction probability estimate of $p = 0.979$ (95% confidence interval [0.971, 0.987]), with the model estimations following the trend of the observed data. These results lend cautious support to the hypothesis that the observed pattern of outbreaks at universities was consistent with that expected from importation of infection from the student intake.

This would imply that outbreaks are more likely when case numbers in the incoming student population are higher (higher n leads to higher outbreak probability $\mathcal{P}$). Similarly, if the extinction probability, p, i.e. the probability of the chain of infection originating from a single introduction dying out, were lower then the overall outbreak probability $\mathcal{P}$ would be higher. Less effective infection control measures or a more transmissible variant might lead to a lower p, but this needs to be investigated further.

2.1.3. Limitations

Factors that we did not take into account in this simple initial analysis and that could be explored further include: the detailed timeline of importations and onward transmissions, the likelihood that an outbreak might be the sum of smaller outbreaks caused by independent introductions, the rate of assimilation of local prevalence in newly arrived students, the impact of heterogeneous university characteristics (such as the number of commuting students), and the impact of heterogeneity in university infection control measures.

In addition, we were limited by the availability of data; ideally the analysis should be repeated with contemporary student numbers and home regions, and with more consistent data on university case numbers.

In light of these limitations, the precise numerical value of the fitted extinction probability should not be interpreted literally. However, the fact that the extinction probability appears to be high suggests that the majority of infection chains die out before sparking an outbreak. This may be partly because COVID-19 is highly overdispersed [9] so that only a small proportion of infections lead to further cases, while many people with the disease do not infect anyone else. It may also reflect effective infection control measures in universities, or that there were fewer incoming infections than assumed in the model, perhaps because students who were unwell may have delayed their return to university,

or because estimates of the prevalence via PCR-testing includes people who are in the late stages of infection and no longer infectious.

2.2. Infection risk in residential student halls

Prior to the resumption of the 2020/2021 academic year there was limited data to relate transmission risk within halls and their households to that estimated for community households. Here, we examine factors predicting risk of infection among students in halls of residence at a single university. We refer to the secondary attack rate (SAR) in a subpopulation (e.g. household, hall of residence) as the probability that a member of the subpopulation is infected following infection of one subpopulation member.

2.2.1. Data

Data on hall capacity for 19 halls managed by the university, and the assignment of rooms within these into households of up to 16 members, were collected prior to the start of term. Stock data on room types for each hall was used to estimate the fraction of students sharing bathroom facilities with at least one other student for each hall and in each household. During term, students were encouraged to report confirmed SARS-CoV-2 infection via a web form, including information about their place of residence, date of test result and subject. Preliminary enrolment data for 2020/2021 by subject and term-time residence were used to estimate the fraction of students in each hall enrolled in the Medical Faculty (as a proxy for students who may be at higher risk of infection due to placements). Approximately half of students reported a room number in addition to identifying their hall of residence, which enabled these reported infections to be grouped into pre-assigned households of known size.

2.2.2. Methods

We tested for predictors of the SAR in a hall using multivariate logistic regression. We included median household size, proportion of students in medical courses, hall size and the proportion of students sharing a bathroom with one or more students as covariates.

We used binomial logistic regression on the binary data indicating the presence of at least one infection in each household to estimate the probability that infection is reported by household size. We estimated the binomial probability of secondary infections in a household. We also considered multivariate logistic regression performed with covariates of household size, time between start of term and date of first reported test in the household, and proportion in the household sharing a bathroom. We aggregated household data across halls and only included reports that were associated with symptomatic SARS-CoV-2 infection, to avoid bias in time between start of term and date of first reported test in the household from asymptomatic testing programmes.

We repeated each multivariate regression while at least one predictor was not significant, dropping the predictor with the lowest *t*-value. We performed the statistical analyses using the general purpose mathematical programming language Matlab [10] (logistical analysis) or statistical data analysis software Genstat [11] (binary logistical analysis).

2.2.3. Results

2.2.3.1. Reported confirmed attack rate by hall

While all covariates listed in table 7 were significant in a univariate analysis (appendix B), only hall size and proportion of students sharing a bathroom were associated with SAR in the final multivariate regression (table 1). We provide the predicted impact of hall capacity and the proportion sharing bathrooms in table 2. This indicated that students in halls where they all share a bathroom with at least one other (shared = 100%) are approximately 50% more likely to become infected than students in halls with all en-suite rooms (shared = 0%). Increasing the hall capacity from 100 to 400 students increased each student's probability of becoming infected by approximately 167%.

Our results—with the caveat that they are subject to any bias in confirming and reporting infection—suggest that infection risk in large residential settings is difficult to mitigate by segmenting students into households, and the risk of living in large residential settings is exacerbated by the use of shared bathrooms. It is possible that our covariates are proxies for other properties of the setting that influence student mixing (e.g. other types of shared spaces, ventilation, etc.). Furthermore, it is likely that effect sizes will vary between settings depending on importation of cases, characteristics of the

Table 1. Coefficients, and associated *p*-value and standard error for the final logistic regression models for the hall SAR.

covariate	coefficient	*p*-value	s.e.
hall size	0.0037	<0.0001	0.00006
proportion shared bathroom	0.4738	<0.0001	0.1166
constant	−3.1466	<0.0001	0.2235

Table 2. Expected impact of increasing hall capacity (size) and proportion of students sharing a bathroom (shared) on the hall SAR (95% CI) from the final multivariate logistic regression in table 1.

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)

local epidemic and local testing facilities, and propensity to adhere to guidance on isolation and mixing restrictions. However, interpreted at face value, our results suggest that only partially filling student residential halls could significantly reduce transmission risk, especially if this is coordinated to reduce shared spaces.

2.2.3.2. Infection risk within hall households

Unsurprisingly, the probability of at least one reported symptomatic infection in a household was significantly correlated with household size (table 3); the expected probability of importation into a household of size 16 was 0.68, approximately double the probability for a household of size 8. Thirty-eight per cent of households reported at least one infection.

Household size does not reach significance in the regression model for household SAR in the univariate or multivariate analysis, consistent with estimates of community household SAR for households from population level data [12]. For the multivariate regression we find that SAR was higher for households with the first reported case earlier in the term (table 3). This has many possible drivers such as changes in local background prevalence, shifts in contact or reporting behaviour, or the impact of local depletion of susceptible individuals owing to immunity or students vacating term-time residences. Our analysis of this dataset does not allow us to distinguish between these possibilities. Multivariate regression also indicated the SAR was positively correlated with the proportion of shared bathrooms in the household. The first reported infection in a hall household occurred six days after the start of term. At this stage of the term our predicted household SAR is 0.09 (95% CI: 0.05–0.16) and 0.21 (95% CI: 0.14–0.30) in households with all en-suite rooms and all rooms with shared bathrooms, respectively.

Although the vast majority of test results within a household were dated within 14 days of the first reported positive, and therefore plausibly epidemiologically linked, we did not have any contact tracing or situational data that could be used to investigate this. We have not estimated overdispersion in the number of secondary household cases which may be relevant [13]. While our estimates of the SAR early in the term are broadly consistent with community household SAR (e.g. [12,14]), the binomial probability of reporting a symptomatic infection given a previously reported symptomatic infection in a household over the entire term is lower: 0.058 (95% CI: 0.043–0.070) or 0.076 (95% CI: 0.064–0.090) considering all reported positive tests. However our data on secondary household infections is incomplete due to missing data on household membership and uncertain propensity to report test results (including any time and household dependence of this). Follow-up testing of household members for markers of historic infection in serum samples is probably required to estimate the full extent of household transmission.

It is highly plausible that not all infections in a household arise from a single imported case. In appendix B, we consider the role of infection within the hall on the household SAR using a simple transmission model that allows for infectious contact between household members and between hall

Table 3. Coefficients, and associated *p*-value and standard error for final regression models for the probability of introduction of SARS-CoV-2 into a household and household SAR.

covariate	coefficient	*p*-value	s.e.
binary logistic regression: probability of infection in household			
household size	0.1623	<0.001	0.0269
constant	1.847	<0.001	0.2510
logistic regression: household SAR			
date of first infection	−0.1485	<0.0001	0.0298
proportion shared bathroom	0.9500	0.0021	0.3091
constant	−1.4028	0.0019	0.4524

members. Results indicate the extent extra-household contacts in the hall may inflate estimates of the SAR; in this model the mean probability of infection due to random contact within the hall is 0.047, whereas the probability of infection from an individual in the same household is 0.091 (see appendix B). In reality, students will also mix with students in other residential settings and with the wider community—we explore evidence for the latter in §2.3.

2.3. Transmission to/from the community: comparison with local age groups

Following a series of large outbreaks among the university student population in the 2020/2021 academic year, a question of interest to both policymakers and the general public was the extent to which these outbreaks affected the wider local communities. This question remains of importance for any future large-scale returns of students to their campuses, and provides insight into the extent to which cluster outbreaks impact nearby populations.

In this section, we examine *spillover*, the impact of outbreaks in student populations on the surrounding communities, by analysing patterns of cases among the student population and the local community. In practice, as student populations are interlinked with the wider community, transmission can be in either direction. In addition to any NPIs in place, and adherence thereto, the existence and strength of any spillover signal will probably depend on factors such as: the magnitude of the student outbreak, the levels of newly reported cases (incidence) in the community at the time of the outbreak, and the proportion of students who originally resided in close geographical proximity to the university.

2.3.1. Data and methods

We used age-stratified positive case data at the lower tier local authority (LTLA) level from a Public Health England (PHE) line list to describe the trends in student-aged case numbers. Our analysis also used cumulative incidence data as reported by the respective universities, or via the University and College Union (UCU) COVID-19 dashboard [8]. Cumulative case counts from both data sources were used as measures of the outbreak sizes. Calculations of these sizes were limited to 10 days past the peak in student-aged cases in order to facilitate comparisons across all LTLAs.

The age-stratified line list data for those aged 18–24 was used as a proxy for 'student cases', with cases among all other age groups being classified as 'community cases'. To facilitate comparison across age groups, we rescaled all quantities by the known populations of each LTLA using data from ONS [15].

We include a sample of LTLAs with a notable proportion of students in table 4 as an illustration of the variability across England. For each LTLA, we examined if, following an outbreak in the student population, (a) there was an appreciable increase in the growth rate of community cases, and (b) if more community cases than expected were recorded in the subsequent 10 days.

The time-varying growth rate in cases was estimated by taking the derivative of a smoother applied to the daily case data. This method, while accounting for overdispersion in the data, also estimated a mean daily incidence (see appendix C for more details).

Upon infection, a host triggers progeny infections following a period termed the generation time. Changes to the community growth rate (a) were regarded as temporally linked with a student outbreak if such significant changes occurred within two generation times (approx. 10 days [17]). Cases in excess of the expected daily incidence were used as a proxy for (b).

Table 4. Properties of each of the considered LTLAs. Local students refers to those students domiciled in the same English region, as obtained from the Higher Education Statistics Agency. The community prevalence was obtained at the regional level from the ONS [16], looking at the transition from 15 September 2020 to 15 October 2020. Multiple return dates arise from those LTLAs which host multiple universities.

LTLA	region of England	local students	ONS prevalence (%)	return dates
Birmingham	West Midlands	52.8%	0.08 → 0.79	21 Sep
Bristol	South West	23.3%	0.08 → 0.30	21 Sep & 5 Oct
Durham	North East	15.4%	0.34 → 1.24	5 Oct
Exeter	South West	32.0%	0.08 → 0.30	14 Sep
Leeds	Yorkshire & The Humber	38.7%	0.25 → 1.51	28 Sep
Manchester	North West	50.0%	0.44 → 1.83	14 Sep & 21 Sep
Newcastle	North East	45.7%	0.34 → 1.24	28 Sep
Nottingham	East Midlands	32.1%	0.13 → 0.69	21 Sep
Oxford	South East	37.1%	0.09 → 0.43	5 Oct
Salford	North West	76.6%	0.44 → 1.83	14 Sep
Sheffield	Yorkshire & The Humber	39.3%	0.25 → 1.51	28 Sep
York	Yorkshire & The Humber	33.3%	0.25 → 1.51	28 Sep

2.3.2. Results

The degree to which the growth rate of community cases changed following a student-aged outbreak varied significantly across the studied LTLAs. A selection of the different observed patterns are included in appendix C.

Figure 3 shows a diverse pattern of spillover, and lack thereof, across different English LTLAs. Unsurprisingly, some of the universities with the largest outbreaks were situated in LTLAs which simultaneously had higher levels of incidence in the community.

Larger outbreaks correlate with a greater degree of spillover, although this effect is more strongly seen when considering cases among 18–24-year-olds in figure 3*b* compared with using reported student outbreak sizes in figure 3*a*. However, there are exceptions to this pattern, and there is not a clear formal relationship between spillover and outbreak size.

Although we consider two separate data sources to gauge campus outbreaks (self-reported or age-stratified), the discussion below uses outbreak sizes from figure 3*a*. At lower levels of community incidence, we observe two scenarios: in the first, a small outbreak with little apparent impact on the community. In the second, an outbreak in excess of 1200 cases with the largest observed impact on the community. In this latter case, the impact was larger in relative terms, but not necessarily in absolute terms (net increase in community cases).

No clear relationship is apparent between the proportion of local students and excess community cases.

Some large outbreaks (in excess of 1750) took place with relatively low levels of excess community cases. It is hypothesized that the asymptomatic testing strategy in place at the university in question may have played a role in this outcome.

2.3.3. Limitations

Student populations are interlinked with the wider community, whereby transmission can occur in either direction. For a given outbreak then, purely from case data it may not be possible to determine whether or not a student population caused or exacerbated an outbreak in the community. Our findings on spillover here are therefore limited to correlations between the growth in positive cases among the student-aged population and the community.

Particular care should be taken when interpreting the relative timings of increased growth rates as done in appendix C, as community cases rose in England during the autumn. In general, our results are limited by the available data, the sample of studied LTLAs, and our chosen indicators of spillover. While the chosen age groups represent those most likely to be students (ages 18–24) and members of

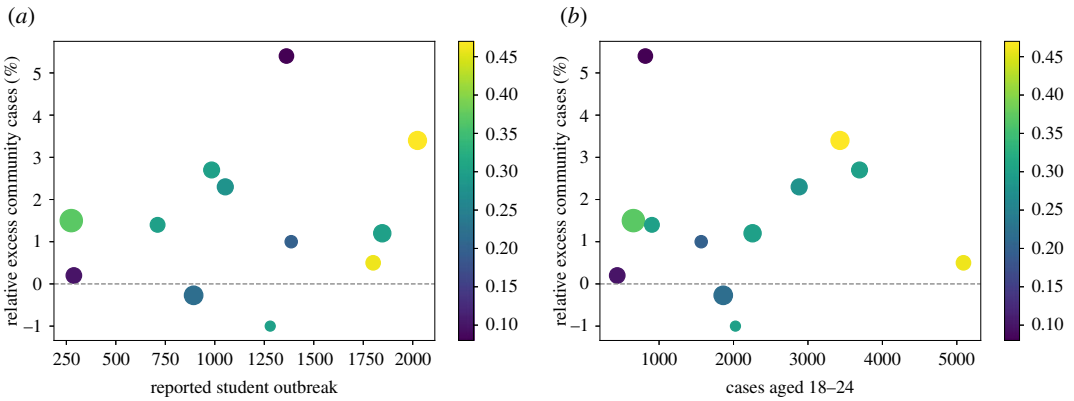

Figure 3. Relative excess of community cases in relation to the reported outbreak sizes across the LTLAs considered in table 4. The sizes of the plot markers scale with the proportion of students attending a university in the same region as their home address. The colours of the markers correspond to the community incidence per 1000 people in each LTLA at the time of peak student-aged cases. These inform the varying levels of community prevalence prior to any student outbreak potentially impacting the community. (*a*) Cumulative university student outbreak sizes up to 10 days past peak incidence, reported by the UCU. (*b*) Cumulative university-aged outbreak sizes from 14 days prior to 10 days past peak incidence, reported by PHE.

the wider community (ages 0–17 and 25+), these age ranges fail to account for older students and those aged 18–24 who are not in higher education (HE).

Since the analysis is based on confirmed cases, our findings are predicated on consistent testing availability and uptake. Significant changes to these over the studied time period may have impacted our conclusions. The values in figure 3 should not be taken as predictive of the impact a student outbreak will have on the wider community. Overall, signals of spillover are not consistent in type (growth rate or excess) or strength across the studied LTLAs. As such, there does not from the data appear to be a simple set of criteria which can be established to determine the risk to the community from a university outbreak.

While our observations suggest that spillover of cases from the university-aged population to the wider community probably does occur, this analysis does not consider transmission settings, e.g. residential, social or educational.

2.4. Transmission to/from the community: spatial patterns

To complement the previous section's spillover analysis based on age-bands, we investigated relationships between the number of cases in areas (middle super output areas, or MSOAs, which are statistical reporting regions in England and Wales typically containing 5000–10 000 people) with a large concentration of students, and areas that are near or far from those student areas.

2.4.1. Data and methods

To estimate the proportion of the population within any given MSOA composed of HE students, we used information on the number of people reporting being students in each MSOA from the 2011 UK census [18], and 2019 mid-year population estimates from the Office for National Statistics [15]. For weekly new case counts by MSOA, we used the public UK government coronavirus data portal [19]. We derived MSOA centroids from the Office for National Statistics geographical data [20].

We defined an MSOA as *high student concentration* if the number of students reported in 2011 was at least 15% of the 2019 population estimate, and *low student concentration* if this figure was below 5%. We classified an MSOA as *near* a high student concentration MSOA if it was not itself a high student concentration MSOA but its centroid was within 2 km of the centroid of such an MSOA, and *far* otherwise. We plotted time series of test-positive cases per population by week for these categories of MSOA in several local authorities.

2.4.2. Results

We find a very mixed picture across different local authorities hosting HE providers across England, and show several examples in figure 4. In particular, we see some signal of spillover in the case of Manchester

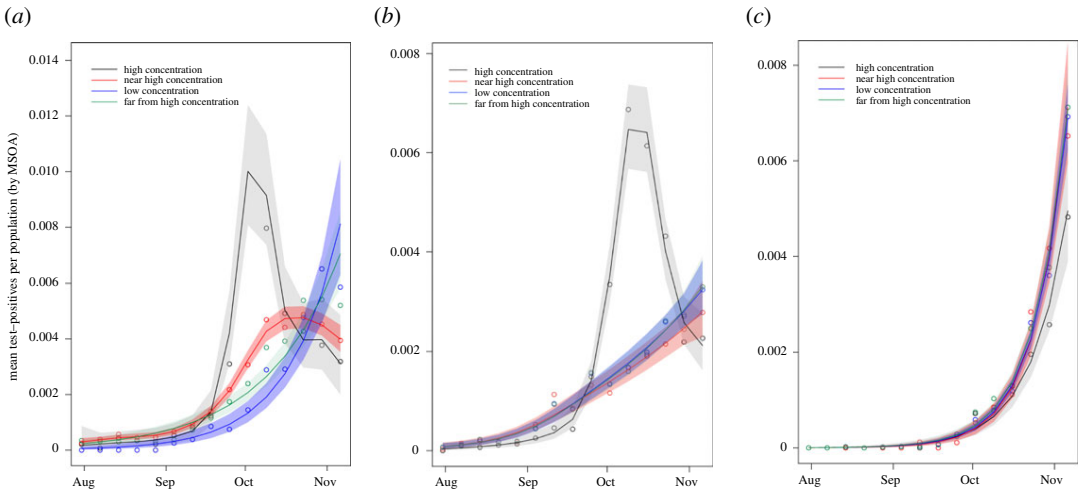

Figure 4. Mean cases (represented as dots) per population in MSOAs categorized as high student concentration (black), near high student concentration (red), low student concentration (blue), and far from high student concentration (green) in each of (*a*) Manchester, (*b*) Birmingham and (*c*) Hull. Lines represent the smoothed weekly mean positive cases per population, shaded to cover the 95% confidence intervals of these estimates (details in appendix C).

(figure 4*a*), where the MSOAs near high-concentration student areas experienced a rise and peak in cases following a rise and peak in high-concentration student areas that is visibly distinct from the pattern for areas that are far from student areas. By contrast, in Birmingham (figure 4*b*) we see a rise and peak in cases in high-concentration student areas, but no distinction between the visible patterns for MSOAs near high-concentration student areas or those further away. In the case of Hull (figure 4*c*), we see no obvious distinction between any of the categories of MSOA. When we combine our age-stratified analyses and these geographical-spread analyses we continue to see a mixed picture: some local authorities have signal of spillover, but some do not. We do not see a consistent pattern across England, probably due to wide variations both in the course of the coronavirus pandemic and the nature of university–community interaction in different local authorities. Considerations such as the severity of imposed NPIs, magnitude of student body, and uptake and efficacy of testing, tracing and quarantining measures probably all influence the overall results, but their individual contributions are not identifiable in this analysis. There is agreement between the age-stratified and geographical-spread analyses of spillover in e.g. Manchester and Birmingham. This supports the robustness of the spillover signals (where observed), and the utility of both methods.

3. Exploratory modelling for future return

During the autumn term in the 2020/2021 academic year, one of the recurring problems that universities encountered was the large number of students that needed to isolate in halls of residence. The isolation was seen as detrimental to the mental health of students, but also the sheer number of isolated students posed logistical problems to the universities. For instance, making sure that students received adequate food packages was a problem at the beginning of the term. It was an ongoing discussion how to reduce the number of students in isolation and to 'flatten' spikes in the number of isolated students to help universities to better deal with these logistical challenges.

The large outbreaks in universities during the first term in the 2020/2021 academic year led to consideration of methods to safely manage the return of students for the second term of the academic year in January 2021. Two constituent components of the initial guidance (published on 2 December 2020) were the staggered return of students and increased usage of rapid tests [21]. Universities were asked to stagger their students returning over a five-week period according to course type. Students in subjects that most required face-to-face interactions, such as medical and veterinary students, were identified to be the first ones to return to campuses. Guidance also stipulated that all students should be offered a SARS-CoV-2 test when they returned to university, helping identify and isolate those who were asymptomatic. The protocol involved two lateral flow tests (LFTs), 3 days apart. In practice, however, this staggered return did not occur as planned in January 2021. Following the imposition of a new

nationwide lockdown on 4 January 2021, there was a prioritization of return of students to face-to-face teaching enrolled on courses that were most important to be delivered in-person in order to support the pipeline of future key workers. All other courses were to continue being delivered online [22].

In this section, we bring together insights from multiple independent models assessing the impact of staggering the return of students to university and mass testing on infection and isolation. The intention of our modelling work was to focus purely on unpicking the epidemiological consequences of staggering student return on SARS-CoV-2 transmission and isolation. We acknowledge there are multiple factors that administrators must consider and there may be operational and/or resource reasons why a staggered return at higher education institutions is desired. These include ensuring that testing capacity is sufficient to meet demand, the monetary costs associated with the intervention (e.g. testing and staffing) and the educational needs of the students. Though the inclusion of these considerations is beyond the scope of our study, they are important constituents of a multi-faceted decision-making process and we provide an expanded discussion in the Conclusions section.

We present work from four independent models that implement a staggered student return, with the view of having multiple approaches (with distinct modelling assumptions) to enhance result robustness and to determine whether consensus findings emerged. We open with two parsimonious model frameworks. The first is used to highlight potential surges in the number of students in isolation upon student return (§3.1). The second presents a transmission model that considers the impact of staggered student return over time (§3.2). The final two models continue the exploration of the dependency of epidemiological outcomes on staggered return policies, with both models incorporating heterogeneity in contact structure and being partly parametrized using data on (different) individual higher education institutions (§3.3).

With respect to mass testing, we consider insights from two network transmission models, each with a differing area of focus. In one analysis we vary the return testing strategy, in conjunction with staggered student return (§3.4). The other considers regular rounds of testing throughout the academic term and the potential implications of a SARS-CoV-2 variant with increased transmissibility, in light of the emergence of the B.1.1.7 SARS-CoV-2 lineage that proliferated rapidly in the UK in late 2020 and early 2021 [23–26] (§3.5).

3.1. Impact of staggering on isolation

To investigate the viability of a staggered return approach, we built a basic discrete event simulation for the return of students to their halls of residence. This individual-based model was designed to investigate the necessary capacity that would be required on campus to isolate incoming students and to establish whether staggering could reduce the overall time that individuals would spend in isolation upon return. In this section, we purely focus upon isolation as a result of a positive test upon return and do not consider spread of infection within the university after students return.

3.1.1. Methods

In the model, each student arrives in their household and is tested immediately. If their test is positive, their household is put into isolation for 10 days. If a particular student is due to arrive in a household that is already isolating, that student is required to wait until the relevant household comes out of isolation before they are allowed to return and have their test.

We investigated four different scenarios: (i) all students return on the same day, (ii) each student returns on a random day in a 14-day interval, (iii) each student returns on a random day in a 28-day interval, and (iv) the students return in three weekend 'pulses'. In these pulses, we assume that 10% of students are in halls already and 40% arrive on the first weekend. The next 30% arrive on the weekend three weeks later and the final 20% arrive on the weekend after that. For the purposes of testing, we treat students that are already in halls the same as the first arrival group. In all cases, we assume that the students that come back at a certain point in time are uniformly distributed over the different households. So, we do not consider effects that appear when, for instance, student housing is organized by programme or year. We note that a fully random distribution of returns over a longer period might be practically infeasible, and assuming that returns are concentrated on, for instance, weekends, is a more plausible assumption.

We simulated these scenarios for cohorts of 1000 students. We varied the household size and the probability of receiving a positive test. The results of these simulations are summarized in table 5,

Table 5. Summary of the staggering simulations. The table shows the average over 100 runs for each combination of household size and fraction of positive tests (3WP: three week pulsed return, *p*: probability of positive test result, *W* + *I*: combined total for either waiting to return or isolating).

household size	*p*	arrival	isolating	waiting	*W* + *I*	peak isolating
10	0.01	3WP	621	99	720	102
		at start	931	0	931	170
		random14	594	218	812	94
		random28	577	129	706	89
	0.02	3WP	1183	186	1369	170
		at start	1812	0	1812	320
		random14	1152	401	1553	178
		random28	1171	255	1426	116
	0.05	3WP	2908	438	3346	303
		at start	4049	0	4049	510
		random14	2868	951	3819	307
		random28	2793	595	3387	256
20	0.01	3WP	1190	187	1378	184
		at start	1806	0	1806	320
		random14	1151	435	1586	183
		random28	1048	250	1299	167
	0.02	3WP	2328	383	2712	279
		at start	3224	0	3224	500
		random14	2275	815	3090	302
		random28	2103	494	2597	228
	0.05	3WP	5512	875	6387	520
		at start	6352	0	6352	780
		random14	5408	1757	7165	552
		random28	5198	1214	6412	512

where we give the total number of days that students need to spend in isolation, need to wait before arriving in their term-time accommodation, and the peak number of students that were in isolation.

We note that, from an organizational perspective for student accommodation, not only the total days spent in isolation is relevant, but also the number of students that are isolated at any given time.

3.1.2. Results

To show the impact we have plotted the average numbers for the different simulations in figures 5–7 for the random return within 14 days, 28 days and the three-pulse return.

We observed that staggering the return of students can have organizational advantages. Under a regime where the fraction of positive tests in the student population is low and household sizes are small (figures 5–7, top left panels), spreading out the return of students can reduce the total number of days that students spend in isolation and also reduce the peak number of students that are isolated on a given day. These advantages diminish or are even reversed if the proportion of positive tests is high (figures 5–7, bottom rows); in that case households are repeatedly put in isolation, which leads to higher peaks and total days in isolation. As can be seen in the case of household sizes of 20 students and positive test probability of 0.05, spreading the return of students over a longer period of time mainly reduces the peak number of isolations and does not contribute significantly to a reduction in the total number of days that students are isolated in these scenarios. We note that for positive test probabilities of $p = 0.02$ and $p = 0.05$, one can expect that a significant number of students

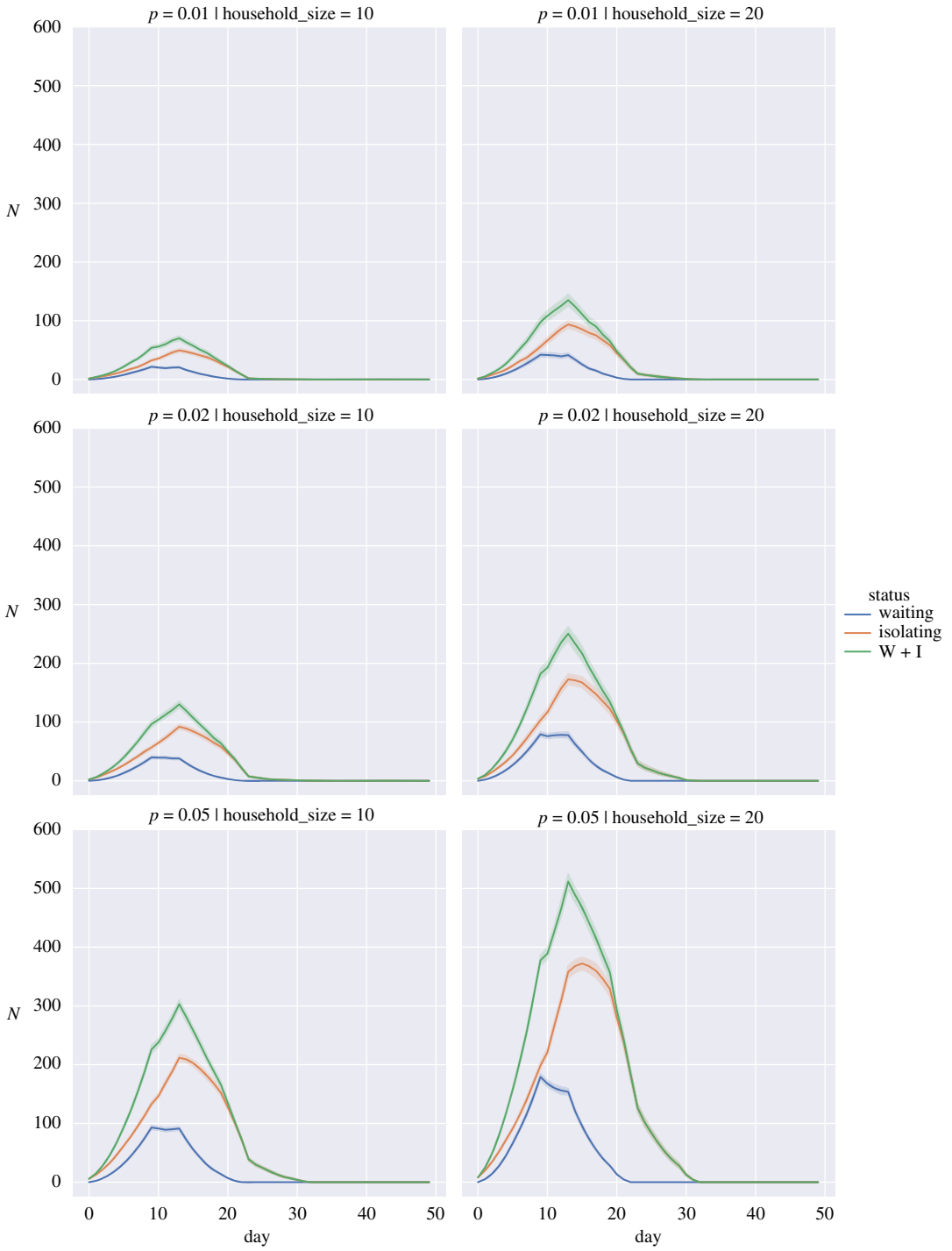

Figure 5. Expected number of students in isolation against time for a return spread over 14 days when the probability of a returning student being infected, p, is 0.01 (top row), 0.02 (middle row) and 0.05 (bottom row) for household sizes of 10 individuals (left column) and 20 individuals (right column). Waiting (blue), Isolating (orange), W + I: Waiting + Isolating (green). Bands show 95% interval computed from 100 simulation runs.

will be impacted by isolation measures in the first weeks after return. Hence, these results suggest that it is important to take this lead time into account when planning in-person teaching activities.

3.2. A simple model for the impact of a staggered student return on incidence

We provide an analysis of the impact of a staggered return of students in three stages, on the transmission dynamics during an academic term.

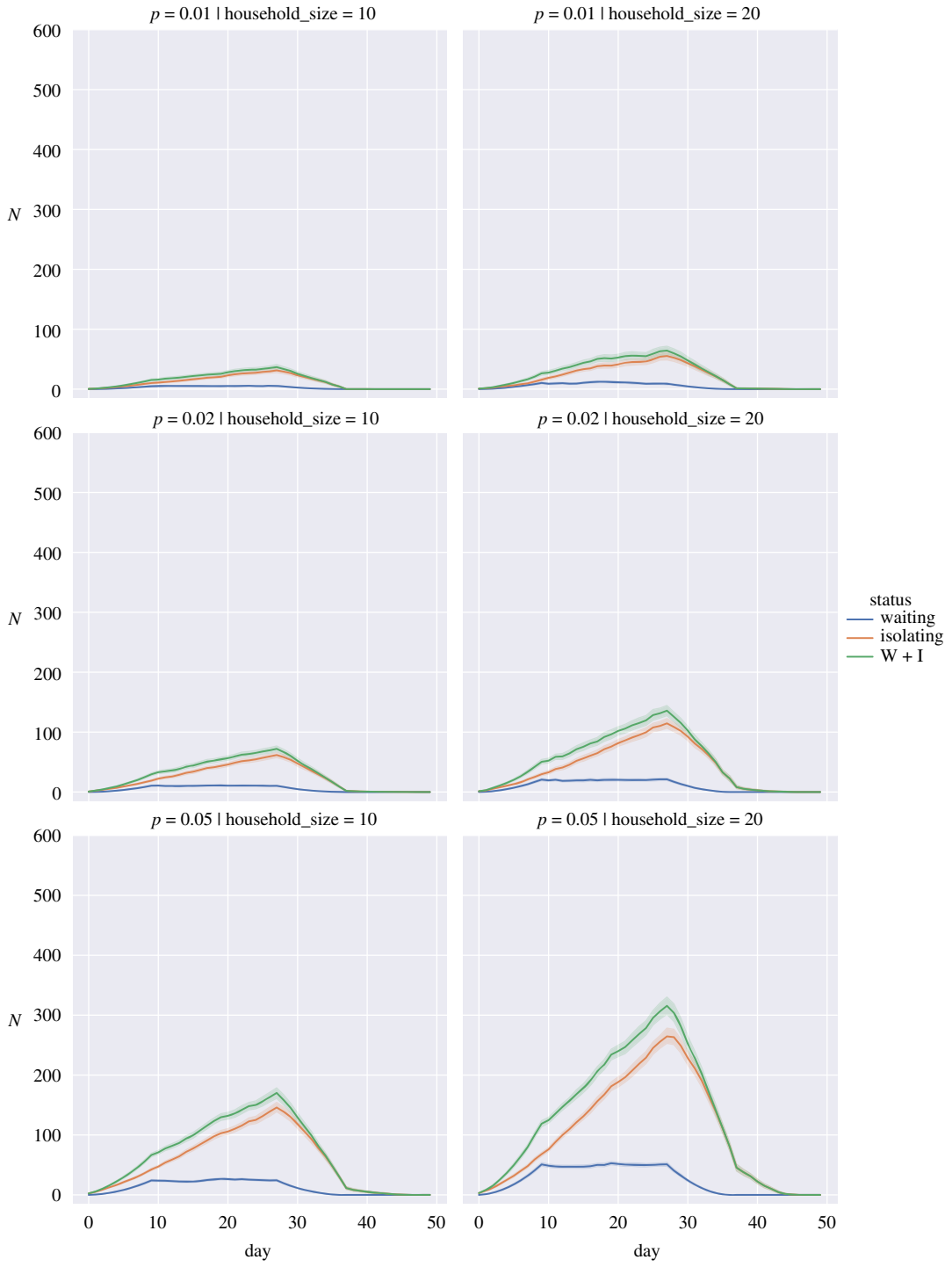

Figure 6. Expected number of students in isolation against time for a return spread over 28 days when the probability of a returning student being infected, *p*, is 0.01 (top row), 0.02 (middle row) and 0.05 (bottom row) for household sizes of 10 individuals (left column) and 20 individuals (right column). Waiting (blue), Isolating (orange), W + I: Waiting + Isolating (green). Bands show 95% interval computed from 100 simulation runs.

3.2.1. Methods

We implement staggered return of students in a simple compartmental transmission model that segments hosts into susceptible (*S*), infectious (*I*) and recovered (*R*) classes, and examine the mean field solutions of this SIR model. We assume that the students return to university in three stages over three weeks. On return, they mix freely with the existing student body and with each other. At each return point, we assume that a fixed proportion of the returnees are infected.

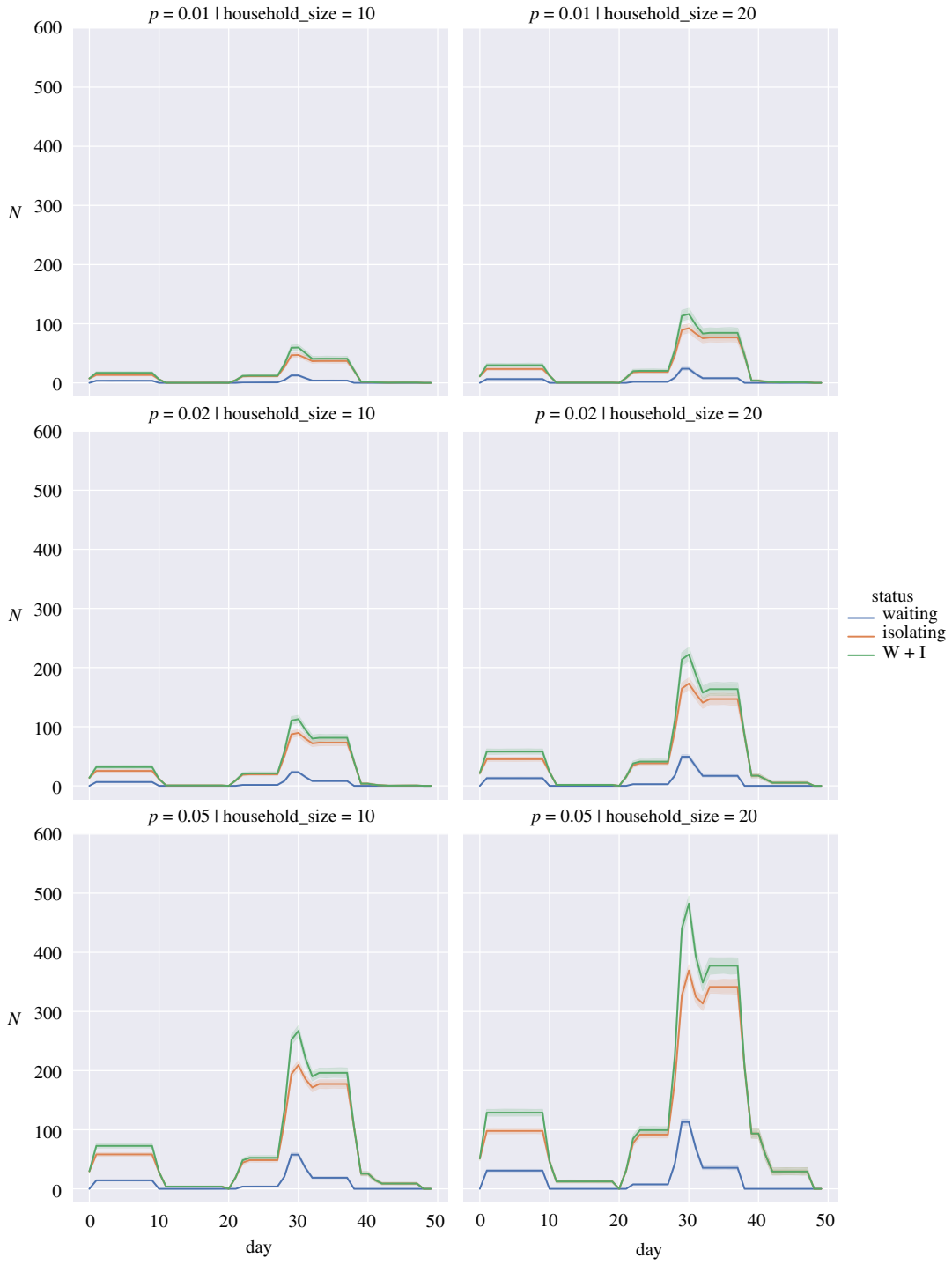

Figure 7. Expected isolations for a three week pulse return when the probability of a returning student being infected, p, is 0.01 (top row), 0.02 (middle row) and 0.05 (bottom row) for household sizes of 10 individuals (left column) and 20 individuals (right column). Waiting (blue), Isolating (orange), W + I: Waiting + Isolating (green). Bands show 95% interval computed from 100 simulation runs.

In our simulation, we took a student body of N students. These returned in groups of $N/3$ in weeks one, two and three, so that the respective student populations in the first three weeks were $N/3$, $2N/3$ and N. Once all of the students return they remain at university for a further eight weeks until the end of an 11-week term.

At each return point, we assume that a fixed proportion, p, of the returnees were infected. In full, when each group of $N/3$ students returned they were assumed to contribute $p\ N/3$ students to the

number of infected students, and $(1-p)N/3$ students to the susceptibles. The resulting SIR model is then given by the following three-level piece-wise model, where $t = [0 \ldots 77]$ was measured in days, $i = 1$ gave the infection dynamics in week 1, $i = 2$ the infection dynamics in week 2, and $i = 3$ the infection dynamics in weeks 3–11:

$$\frac{\mathrm{d}S_i}{\mathrm{d}t} = -\beta\frac{IS}{N_i},$$

$$\frac{\mathrm{d}I_i}{\mathrm{d}t} = \beta\frac{IS}{N_i} - \gamma I$$

and

$$\frac{\mathrm{d}R_i}{\mathrm{d}t} = \gamma I.$$

In each of the three stages, the population values were: $N_1 = N/3$, $N_2 = 2N/3$ and $N_3 = N$.

To simulate the staggered returns, we took the values of S and I at the start of the first, second and third weeks to be the following, noting that the values of S and I then jump at the start of each week (as can be seen in the figures):

$$S_1(0) = (1-p)N/3, \quad I_1(0) = \frac{pN}{3},$$

$$S_2(7) = S_1(7) + (1-p)N/3, \quad I_2(7) = I_1(7) + \frac{pN}{3}$$

and

$$S_3(14) = S_2(14) + (1-p)N/3, \quad I_3(14) = I_2(14) + \frac{pN}{3}.$$

For simulation examples, we used a population size of $N = 1000$ and considered three scenarios with different values of the prevalence (p) and transmissibility (β): (i) $p = 0.10$, $\beta = 0.18$; (ii) $p = 0.02$, $\beta = 0.18$; (iii) $p = 0.02$, $\beta = 0.30$. In all scenarios, we fixed the recovery rate $\gamma = 0.072$. We also compared the results of the 'staggering' model with that of an unstaggered model (with the same parameter values) in which all of the N students returned at the start of term.

3.2.2. Results

The corresponding reproduction numbers R for the three scenarios are initially: (i) $R = 2.5$, (ii) $R = 2.45$ and (iii) $R = 4.08$.

In the absence of all other controls, and across all three considered scenarios, we observed that staggering can slightly reduce and slightly delay the size of the infection peak in the short term (figures 8–10). However, over the course of the 11-week term the reductions in the overall attack rate were minor, particularly for infections with high transmissibility (figure 10).

While based on relatively simple assumptions, these results are intuitive. In conclusion (i) a staggered return could delay and reduce the outbreak peak, however, (ii) without other controls, staggering will not much reduce the overall attack rate over the course of an academic term.

3.3. Structured models assessing the impact of a staggered student return

The formerly presented parsimonious models provide guiding principles on the potential impact of staggering on infection throughout the course of an academic term and isolation upon return. In this section, we build on the prior work by investigating the role of staggered student return on epidemiological outcomes using models incorporating additional layers of complexity. In contrast to the compartmental model in §3.2, these models are simulated probabilistically to explore the random/stochastic variation in outcomes. Specifically, we used two models of transmission dynamics for SARS-CoV-2 in a university setting, each using a different model conceptualization: (i) a stochastic compartmental model [6] and (ii) a network-based model [27]. Both transmission models assume that upon exposure hosts enter a period of latent infection during which they are not infectious, then hosts may remain asymptomatic throughout their infection (asymptomatic cases) or transition through presymptomatic and symptomatic stages of infection. Mass asymptomatic testing may detect both presymptomatically infected hosts and asymptomatic cases. Note that both models assumed that individuals did not 'compensate' by replacing contacts that were unable to occur (due to the expected contact being in isolation or not having yet returned to the university setting).

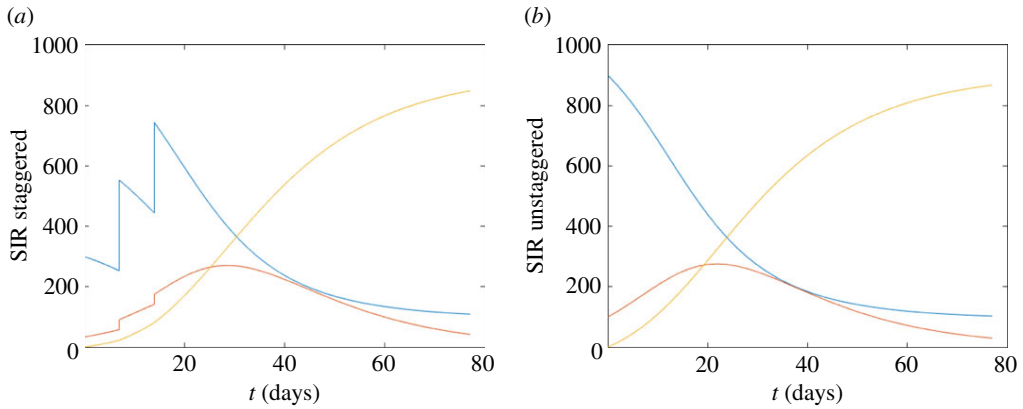

Figure 8. Staggered/unstaggered return temporal profiles. (*a*) Staggered return (three stages in the first three weeks) in the first three weeks, taking $N = 1000$ and $\beta = 0.18$, $p = 0.1$, $\gamma = 0.072$, with an initial value of $R = 2.25$. (*b*) Unstaggered return. In each figure, we show S (blue), I (red) and R (yellow).

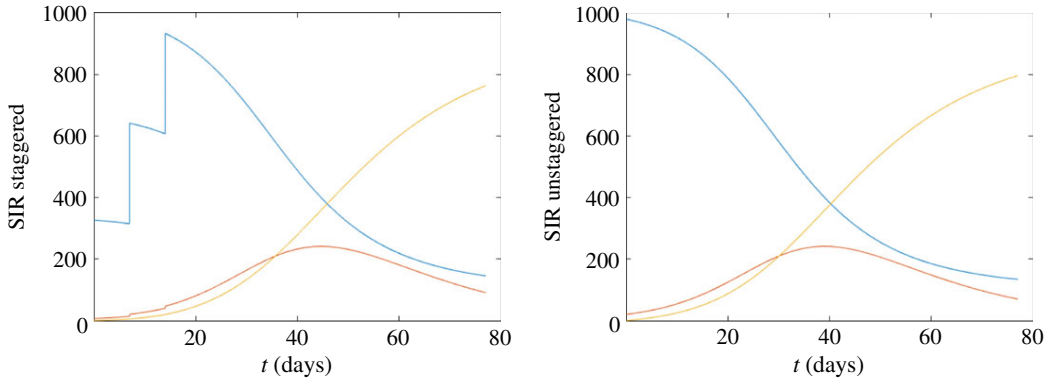

Figure 9. Staggered/unstaggered return temporal profiles. $\beta = 0.18$, $p = 0.02$, initial $R = 2.45$.

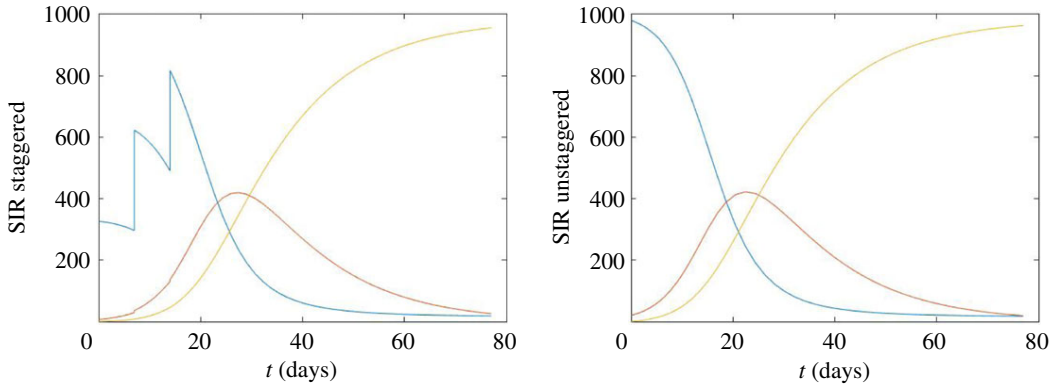

Figure 10. Staggered/unstaggered return temporal profiles. $\beta = 0.3$, $p = 0.02$, initial $R = 4.08$.

3.3.1. Methods

3.3.1.1. Stochastic compartmental model summary

The stochastic compartmental model included realistic mixing patterns for students based on student responses to the Social Contact Survey conducted in 2010 [28,29]. These contact matrices entailed 160 groups based on school (department) and year of study, with contacts stratified into household, study and random contacts. We calibrated the disease compartments to estimations made at the start of the 2020/2021 academic year in the absence of controls, returning an R of approximately 3 (for calibration we assumed asymptomatic cases were 50% less infectious than symptomatic cases). Further model details, including descriptions of the remaining assumptions underpinning the model, may be found in [6].

For our analysis here, we fixed the mean probability of a case being asymptomatic at 75% and the relative infectiousness of an asymptomatic we varied between 0 and 1. It was assumed that the university would operate within Public Health England guidelines and therefore that symptomatic cases would be tested and self-isolate within 48 h. Students in large halls of residence were assumed to be restricted to households of 24 individuals, reflecting actions taken by universities in the 2020/2021 academic year. We did not include the impact of contact tracing, social distancing or the use of face coverings. We used a student population size of 28 000. The number of infected students at the start of term was estimated using home location and incidence as of July 2020 as described in [6] using an anonymized extract of student data for a specific university relating to the 2019/2020 academic year. The study complied with the University data protection policy for research studies [30]. Each scenario was run for a simulated 300 days, with 10 replicates per scenario.

The model is coded in R and C++ and available at https://github.com/ellen-is/unimodel.

3.3.1.2. Network model summary

Our network model framework represents interactions between students within a university population in different settings (household, study cohort, organized societies and sports clubs, other social). We ran an epidemic process on this network, for the virus SARS-CoV-2. The model includes isolation and contact tracing. We adopted a pessimistic approach by assuming a comparable amount of mixing to pre-pandemic circumstances, and did not include any reduction in the risk of transmission occurring over contacts due to social distancing and/or the use of face coverings.

Specifically, we assumed students had contact with all household members each day. We sampled the number of non-household contacts from distributions fit to data informed by student responses to the Social Contact Survey conducted in 2010 [28,29], with stratification according to the level of study (undergraduate or postgraduate). For this analysis, we then applied the following two contact pattern changes to all but the baseline (no intervention) scenario: (i) society contacts did not occur (transmission risk therefore zero), assuming that all meetings would take place online; (ii) for on-campus resident students, we assumed no contacts within the broader accommodation unit of the same floor or block of residence (thus outside the immediate household).

In all simulations, we had an overall student population of 25 000, with 7155 students resident on-campus and the remainder off-campus. Each simulation run had a duration of 11 weeks, encompassing both a 10-week academic term and the week prior to its commencement.

We initialized latent, infectious (asymptomatic, presymptomatic and symptomatic) and recovered individuals using estimates for 2 January 2021 from the University of Warwick SARS-CoV-2 transmission model [31], based on fits from 29 November 2020 and assuming no change to adherence in NPIs.

For each parameter configuration, we ran 1000 simulations, amalgamating 50 batches of 20 replicates; each batch of 20 replicates was obtained using a distinct network realization. We performed the model simulations in Julia v. 1.4–1.5. The data and science surrounding the SARS-CoV-2 infection is fast moving. This piece of sub-analysis was originally undertaken in December 2020, with our intent being for this work to provide a record of the state of our modelling at that time. For a full description of the network model and noted limitations of the methodology, see [27]. We summarize in appendix D other changes made from the base model to carry out this analysis. Distributions of outcome measures are visualized using violin plots which capture the smoothed probability density of a set of numeric values [32].

3.3.1.3. Staggered return strategies

We assessed four strategies for the return of students for the academic term (figure 11) using the stochastic compartmental model and the network-based model. Note that, across all considered strategies, a proportion of the student population was considered to be resident in university accommodation between academic terms.

The four strategies were as follows: (i) no stagger—for students not resident in university accommodation over the vacation, they return on day 1; all students entered the return test procedure on day 1 (we acknowledge that in practice there would be logistic difficulties associated with such a strategy); (ii) 14-day spread—each student is allocated their day to return to university (if applicable) and they begin the return testing procedure between days 1 and 14 (sampled according to a uniform distribution); (iii) 28-day spread—similar to the 14-day spread strategy, except the applicable range spans days 1 to 28; and (iv) three-weekend pulse (by course)—fractions of the student population return on designated weekends based on level and course of study. In the stochastic compartmental model, for the

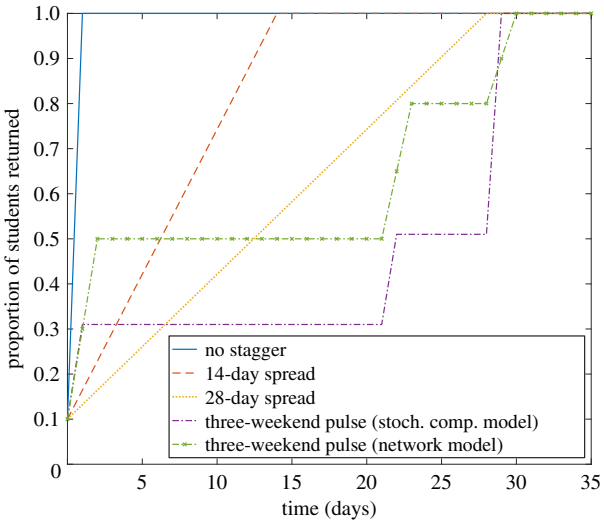

Figure 11. Staggered return temporal profiles. We considered four student return patterns: no stagger (blue solid line); return spread over 14 days (orange dashed line); return spread over 28 days (yellow dotted line); three-weekend pulsed return (by course), as used in the stochastic compartmental model (purple dot-dash line); three-weekend pulsed return, as used in the network model (green dot-dash line, cross markers). For this depiction, we present proportion returned with respect to time when assuming 10% of all students were resident in their university accommodation between academic terms.

three-weekend pulse, on day 1 of the simulation we assumed that all vital medical, dental and veterinary students enrolled in courses (as provided by the University of Bristol [33]) were present, as well as 20% of students in all other schools, giving 31% of students present at university in total. This first group of students was chosen because they were studying on the courses that were allowed to return when universities were closed in January 2021 and at this time it was estimated that 20% of students who were not enrolled on these courses still chose to return. On day 22 of the simulation, all other courses with important practical elements return to university, giving a total of 51% of students present at university. On day 29, all remaining students return to university. For the network model, we set the groupings (and the associated proportion of students returned) for the three-weekend pulse as a variation on the University of Warwick plan for staggering student return [34].

3.3.1.4. Testing protocols

We also included a testing protocol that adherent students engaged with upon return to university. In the stochastic compartmental model, we considered two scenarios: (i) no testing on student return, and (ii) testing of all non-symptomatics. We assumed the tests detect half of true positives (50% sensitivity) and do not generate false-positive results (100% specificity).

In the network model, we assumed adherent students underwent two LFTs, 3 days apart, with isolation between tests (for details on test sensitivity and specificity, see appendix D). For each strategy for student return, we sampled the proportion of students that were adherent to isolation from zero compliance (value 0) to full compliance (value 1) in increments of 0.1. We assumed an identical adherence to isolation restrictions independent of the cause (presence of symptoms, household member displaying symptoms, or identified as a close contact of an infected by contact tracing). Additionally, we assumed those that would engage with isolation measures would also engage with contact tracing.

3.3.2. Results

3.3.2.1. Stochastic compartmental model results

We first present our findings from simulations carried out with the stochastic compartmental model. The collection of simulations that we present here give an indication of what the impact of staggering and testing might have been at the start of the 2020/2021 academic year, if this had taken place. The model parameters do not change based on events that have happened since this time, including vacation periods, and consequently the results are to be interpreted qualitatively if used to make predictions about future scenarios.

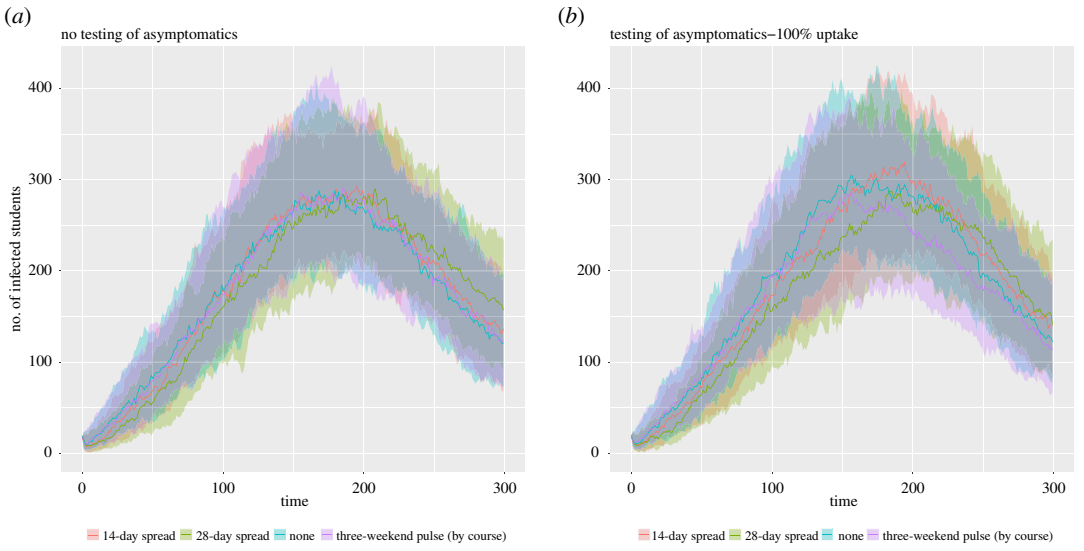

Figure 12. Epidemiological outcomes among a student population given differing staggered return strategies to university using a stochastic compartmental model. Outputs are summarized from 10 simulations, with the lines representing the median number of symptomatic and asymptomatic students and the shaded areas showing the 2.5th and 97.5th percentiles. We display distributions corresponding to: (*a*) no testing of asymptomatics upon student return, (*b*) all asymptomatics are tested.

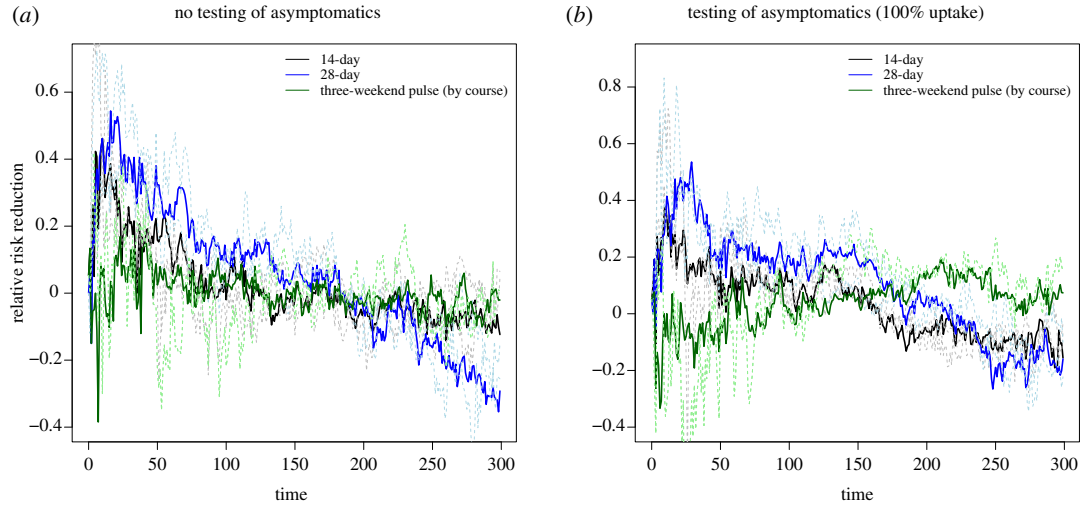

Figure 13. Epidemiological outcomes among a student population given differing staggered return strategies to university compared with a strategy where staggering is not used, using a stochastic compartmental model. Outputs are summarized from 10 simulations, with the continuous lines representing the median number of symptomatic and asymptomatic students and the dashed lines corresponding to the 2.5th and 97.5th percentiles. We display distributions corresponding to: (*a*) no testing of asymptomatics upon student return, (*b*) all asymptomatics are tested.

We observed a similar overall case burden across all considered staggering strategies. Given high adherence to control, similar temporal trends were observed regardless of the testing strategy used (figure 12). Relative to an unstaggered return, there was lower prevalence in the early phase paired with higher prevalence in the late phase for the 14-day and 28-day strategies, with these relationships being consistent across the collection of test upon student return protocols (figure 13).

3.3.2.2. Network model results

For the independent analysis performed using the network model, on account of the inherent uncertainty in several parameters of the model and assumptions made regarding contact patterns, we once more

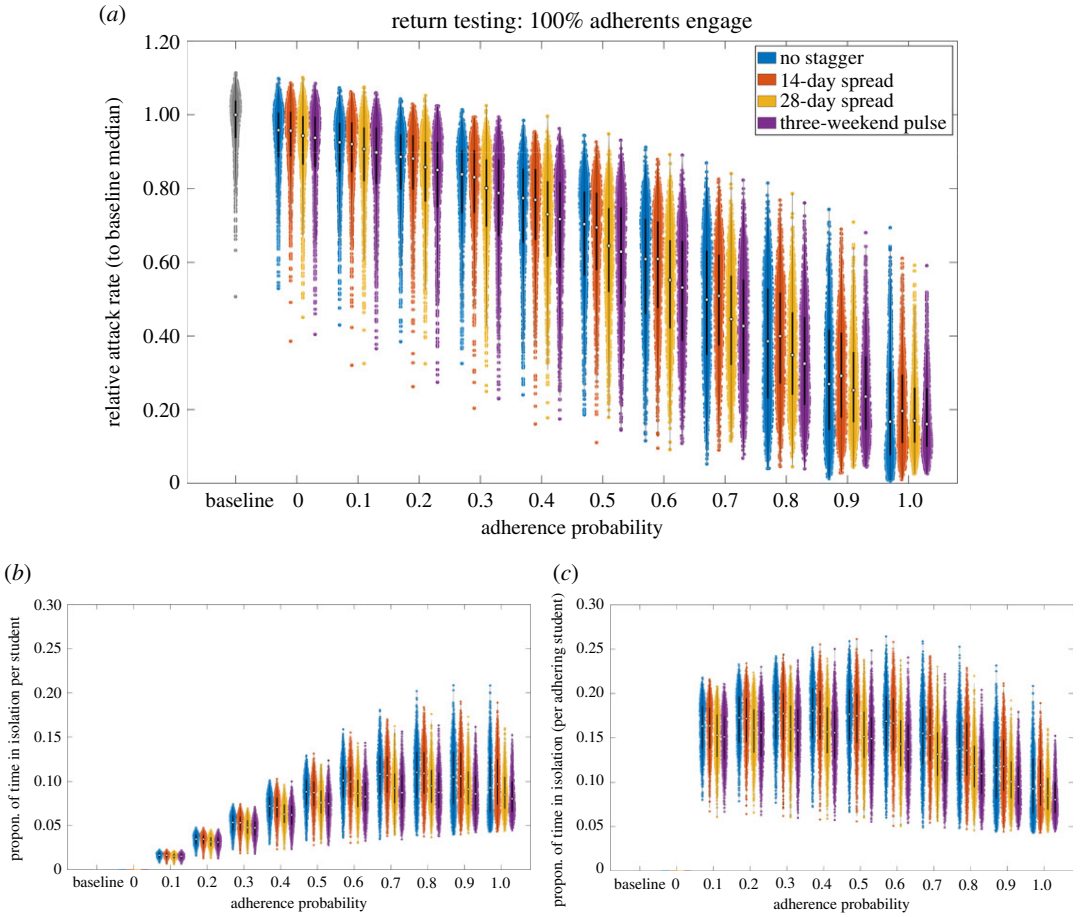

Figure 14. Epidemiological outcomes among a student population given differing staggered return strategies to university. Outputs summarized from 1000 simulations (with 20 runs per network, for 50 network realizations) for various levels of adherence to NPIs. We considered four strategies: no stagger (blue violin plots); return spread over 14 days (orange violin plots); return spread over 28 days (yellow violin plots); three-weekend pulsed return (purple violin plots). We assumed 100% of adherents engage with return testing. We display distributions corresponding to: (*a*) relative attack rate, compared with the baseline scenario; (*b*) time spent in isolation per student; (*c*) time spent in isolation per adherent student. The white markers denote medians and solid black lines span the 25th to 75th percentiles.

focus on qualitative comparisons across the simulated scenarios (as done with the stochastic compartmental model). We first note that, compared with the baseline scenario, the scenario with reductions in contacts via organized societies and dynamic on-campus accommodation contacts (represented by adherence probability 0.0 in figure 14) produced a shift downwards in the obtained distributions of relative attack rate (medians of 0.93–0.96 across the four staggering strategies).

Comparing attack rate across staggering strategies for a fixed adherence level, in concordance with the stochastic compartment model we found a minimal impact on the attack rate over the course of the academic term. Furthermore, we determined adherence to isolation guidance and following test and trace procedures as crucial in reducing the overall case burden within the student population (figure 14*a*).

Assessing the potential impact of staggered return strategies on the amount of time students may be required to isolate, for a fixed adherence level there were no substantial differences between the strategies we considered (figure 14*b*,*c*). Inspecting a measure of time spent in isolation for any given student, we observe an initial increase with adherence level, peaking when roughly 70–80% of students are adherent, before declining as it approaches all students being adherent (figure 14*b*). A collective response (high adherence) reduced the time each adherent student was estimated to spend in isolation, compared with a scenario of moderate adherence among the student population (figure 14*c*).

In the absence of other interventions, staggering slightly reduces and delays the size of the peak, though the long-term impact is minimal (figure 15*a*). For strong adherence to interventions, temporal trends were found to be broadly similar regardless of the staggering strategy used (figure 15*b*), in

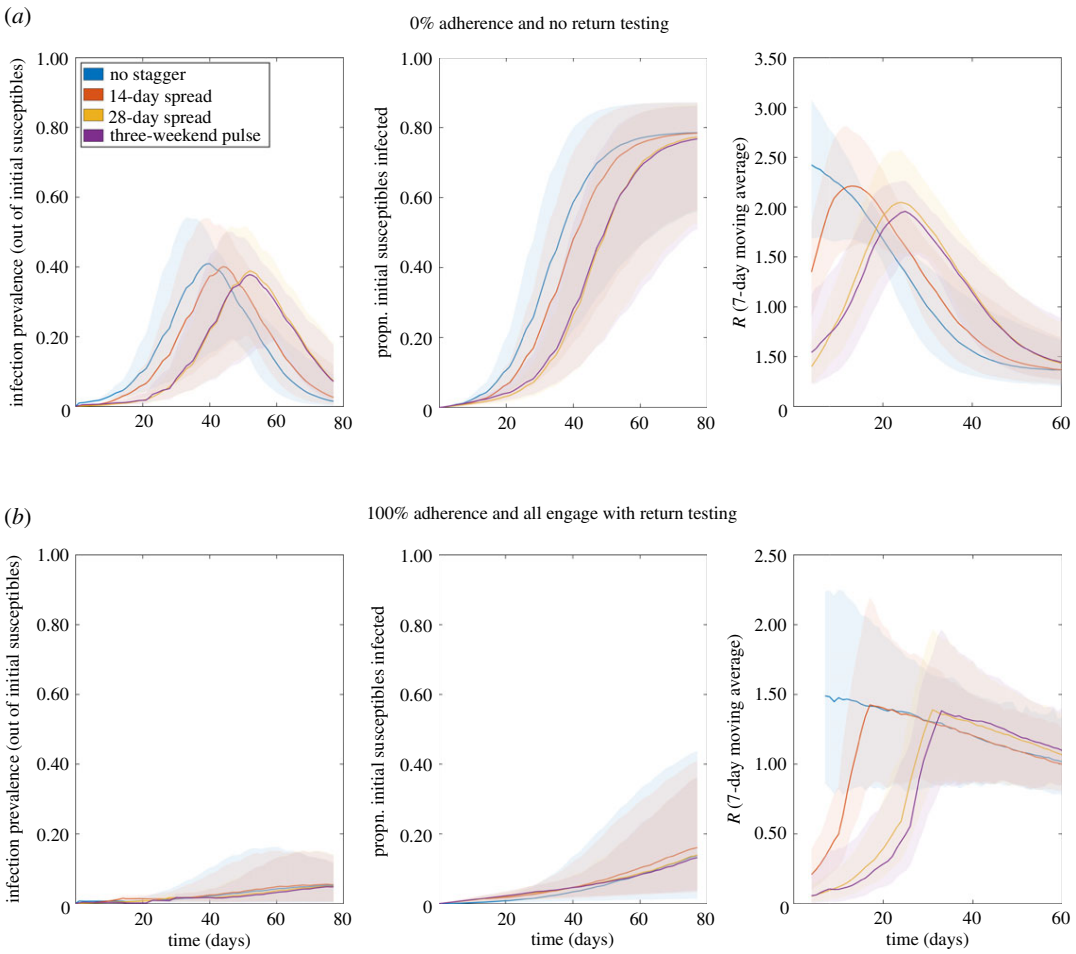

Figure 15. Temporal profiles of epidemiological measures over the academic term under differing return patterns. Outputs produced from 1000 simulations (with 20 runs per network, for 50 network realizations) for four return patterns: no stagger (blue); return spread over 14 days (orange); return spread over 28 days (yellow); three-weekend pulsed return (purple). Solid lines depict the median profile and shaded regions the 95% prediction interval. Panels from left to right display infection prevalence, cumulative proportion of initial susceptibles infected, and 7-day averaged *R*, respectively. (*a*) No return testing; (*b*) return testing with all adherents participating.

agreement with the temporal trends observed from the stochastic compartmental model projections (figure 12).

3.4. Testing on return

Using the network model described in §3.3, we modelled implementation of a testing protocol that students would be advised to complete before attending face-to-face teaching.

3.4.1. Methods

To investigate the sensitivity of staggered returns to alternative test-on-return strategies, using a fixed high level of adherence (90%), we investigated four protocols (table 6). Test protocol A: two LFTs, 3 days apart, with isolation between tests (the default assumption); test protocol B: single LFT; test protocol C: two LFTs, 3 days apart, with no isolation between tests; test protocol D: single PCR with isolation until test result received (2-day delay), leaving isolation upon a negative test result.

3.4.2. Results

Given high adherence to interventions and engagement with rapid testing, the inclusion of a second LFT and isolation between the LFTs gives minor reductions in attack rate (comparing A–D in figure 16). We

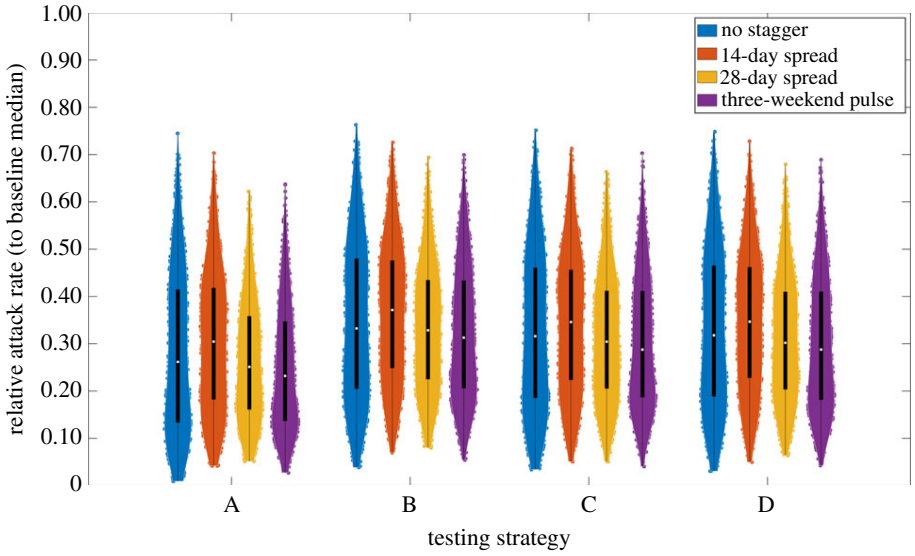

Figure 16. Relative attack rate distributions under different test before return to study procedures, in combination with strategies for staggered student return. Assumed 90% adhere to isolation, test and trace guidance. For test strategies using two LFTs, the two tests were spaced 3 days apart. We considered four student return patterns: no stagger (blue violin plots); return spread over 14 days (orange violin plots); return spread over 28 days (yellow violin plots); three-weekend pulsed return (purple violin plots). The white markers denote medians and solid black lines span the 25th to 75th percentiles.

Table 6. Overview of the return test protocols. Cells containing an 'X' denote the element being a part of the return test protocol. LFT 1 and LFT 2 correspond to a first and second LFT respectively. Conditioned on the plan including individuals undergoing two LFTs, 'isolate between tests' reflects whether isolation should occur between the two LFTs.

testing strategy	LFT 1	LFT 2	isolate between tests
A	X	X	X
B	X		
C	X	X	
D	single PCR test		

found comparable attack rate distributions across our four (previously introduced) staggering strategies for student return to university (comparing between colours in figure 16).

3.5. Testing during term

To build on our investigation of testing on arrival, we simulated the impact of an asymptomatic testing system in use throughout the term, assuming the presence of a more transmissible SARS-CoV-2 variant. This scenario was considered in response to the emergence of the B.1.1.7 variant in the UK, which began to become widespread from November 2020.

3.5.1. Methods

We used a layered network model of contact between 15 000 simulated students, with one layer of household contacts and one of other-group contacts intended to simulate all out-of-household contact. Individuals could be infected by either household or non-household contacts. Infected individuals progressed through disease states via a stochastic compartmental model including a latent period, various infectious states (presymptomatic, asymptomatic or symptomatic) and recovery resulting in immunity (which we assumed did not wane).

We investigated five during-term asymptomatic testing scenarios, in which individuals were tested at random with probability 1/3, 1/7, 1/10 or 1/14 per day (to simulate testing every 3, 7, 10 or 14 days,

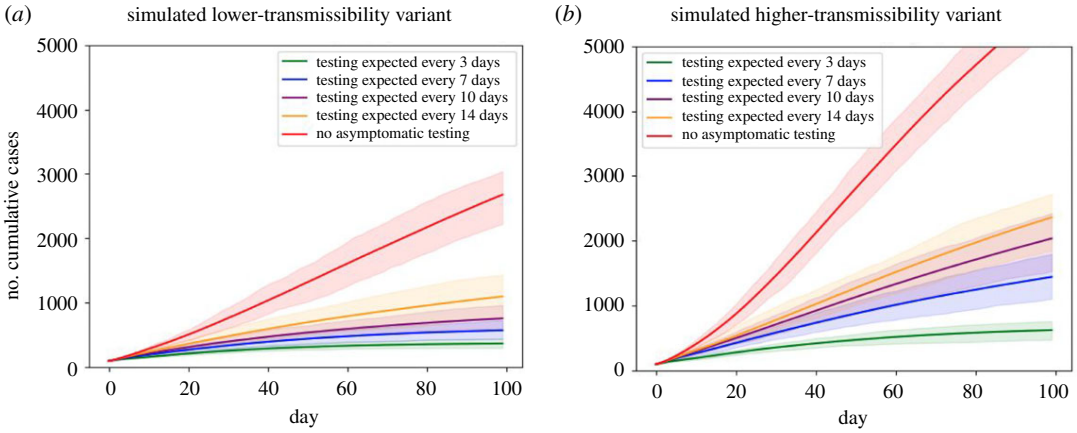

Figure 17. Temporal profiles of cumulative case counts for a simulated population of 15 000 students under differing during-term asymptomatic screening scenarios. We present two scenarios for variant transmissibility: (*a*) lower-transmissibility variant; (*b*) higher-transmissibility variant (1.5 times more transmissible than the lower-transmissibility variant). Output produced from 100 runs of each scenario, with a new network generated for each replicate; envelopes show 95% of model runs and solid lines show mean values. Asymptomatic screening scenarios considered are: no asymptomatic testing (red), each person randomly tested with probability 1/14 (yellow), 1/10 (purple), 1/7 (blue), or 1/3 (green) per day, to simulate testing approximately every 14, 10, 7 or 3 days, respectively. *Note that this model has many limitations and should be interpreted mainly qualitatively. See main text for a listing of some limitations.*

respectively), or not at all. In all scenarios, symptomatic individuals are assumed to be tested immediately upon developing symptoms. Upon a positive test, the entire household isolates for 14 days. Simplifying assumptions included perfect and rapid testing and perfect adherence to testing and isolation. We assumed 50% of non-household contacts to be traced and isolated.

We first ran these scenarios with a lower-transmissibility variant intended to plausibly simulate the variant of SARS-CoV-2 circulating in universities in the UK in autumn 2020. We then considered a 1.5 times more transmissible variant, intended to simulate a potentially more transmissible variant such as B.1.1.7.

We initialized each simulation with 100 infectious individuals, and ran the model for 100 timesteps (notionally days). For each scenario, we performed 100 replicates, each run on a newly generated network. Importantly, we chose the particular parameters for this model for a combination of plausibility and simplicity, and some are not well-founded in any particular dataset. Details of the model, parameter choices and limitations are available in appendix E.

3.5.2. Results

We plot the number of cumulative cases as a time series under the differing testing scenarios for the two variants in figure 17. In general, more frequent asymptomatic screening better controls cases, with the scenario with no asymptomatic screening seeing the largest number of cases. While cases were contained to a mean of fewer than 1200 in all scenarios with asymptomatic screening in the less transmissible setting, this was only achieved by the most frequent testing scenario in the more transmissible setting.

3.5.3. Limitations

This model has many simplifying assumptions and the absolute numbers it produces should not be considered in isolation or as an absolute prediction. Some of these limitations include: perfect adherence to testing and isolation, no vaccination nor prior immunity, no reactive interventions during the course of the simulation, and a speculative network contact structure that *has not been trained from data* but is instead simply a plausible simple structure. In addition, the model did not include a reduction in the risk of transmission occurring over contacts due to face covering use or social distancing; however, other work [35] suggests that if such measures are in place in a university setting

and/or if there are moderate levels of immunity, the impact of testing is less prominent, highlighting the importance of considering testing in the context of other measures.

4. Conclusion

The mass migration of students at the beginning and end of academic terms, their unique living arrangements during term time and unique patterns of social mixing, make them an important population for the spread of infectious respiratory illnesses. Despite this, prior to the COVID-19 pandemic, there was little data collected on outbreaks of infectious disease at universities (although one such dataset collected between October 2007 and mid-February 2008 has now been analysed in 2021 [36]) and university students were an understudied population. Therefore, at the start of the pandemic, there was a limited evidence base to support policy decisions around universities. Our study brings together expertise from multiple research groups and presents results from multiple statistical and modelling analyses and provides new understanding on infectious disease outbreaks at universities and how these could be mitigated.

An important finding of our study is that adherence to NPIs is likely to have more impact than staggering the return of students to university. Survey data suggest that in the autumn term of 2020, students generally did have high adherence to NPIs; an Office for National Statistics (ONS) survey found high adherence (90%) to social distancing across multiple universities [37]. In addition, a survey of University of Bristol students found that 99% of students self-isolated after testing positive for COVID-19 and the majority of survey participants reported low contact numbers [38]. However, there was heterogeneity in adherence, with some students reporting many contacts and with only 61% of students with cardinal COVID-19 symptoms self-isolating [38]. In future, it will be important that students maintain their high levels of adherence and to ensure they have sufficient resources to allow them to do so.

Several of the scenarios presented here have considered the frequency of asymptomatic screening at universities. This has been explored in other modelling studies, for example [39] found that monthly screening can reduce cumulative incidence by 59% and weekly screening by 87%. We found that increasing the frequency of asymptomatic screening is likely to be important in the presence of a more transmissible SARS-CoV-2 variant, with cases only being able to be maintained below 1200 (mean cumulative over 100 days) when testing occurs every 3 days (in a population of 25 000). This finding corroborates a study that used an agent-based model to simulate COVID-19 transmission at the University of California San Diego, where larger outbreaks resulted in a maximum outbreak size of 158 when asymptomatic screening occurred monthly and 7 when it occurred twice weekly [40], but with a much lower impact seen on the average outbreak size when increasing from monthly to twice weekly testing, ranging from 1.9 to 1.1, respectively. Brooks Pollock *et al.* [6] also found that mass testing was more effective for higher values of the reproduction number. This highlights the importance of reassessing control measures under different variants.

We have focused here on COVID-19 risks and mitigation strategies for when students return to university and during the university term itself; however, we have covered little on the risk of transmission from infected students to private homes at the end of term. Previous modelling work suggests that in an unvaccinated population, an infectious student would on average generate just less than one secondary within-household infection, but this is dependent on the prevalence in the student population at the time of departure [41]. Although it is expected that vaccination will reduce the impact of students returning to private homes at the end of term, the UK vaccination programme is ongoing, and there are particular spatial areas and demographic groups where low uptake is expected [42], suggesting that this still may be an important question to consider in future.

Our analyses and discussions have highlighted several areas that we recommend for further attention. These include building a better understanding of determinants of adherence, including attributes that may place subpopulations at higher risk (e.g. students in part-time employment). Given the need for rapid turnaround of our analyses, a persistent challenge is the ability to access data in a timely manner and ensuring any barriers to data access have a purpose and are necessary. One mechanism for addressing this data availability issue may be a centralized nationwide student testing data resource, which could serve as a hub for anonymized student testing data that documents institution and attributes such as type of accommodation.

We recognize there are prominent factors that we have not addressed here, as we have focused directly on transmission dynamics, yet should be considered while viewing our results in a broader context. One

important future research direction is to consider the non-COVID impact of intervention measures. The majority of work to date on COVID-19 has focused upon developing intervention policies that seek to minimize the overall number of cases, hospital admissions or deaths. However, it is important to acknowledge that any control policy that may reduce transmission also has an impact in terms of monetary cost, non-COVID health, mental health and well-being. An extension to this work could focus upon assessing the direct monetary cost of intervention policies as well as the logistical and operational constraints associated with such policies [43,44]. For example, to sustain a regular testing regime at universities under financial, logistical, or structural constraints, mathematical modelling suggests that pooling RT-qPCR testing may be a cost-effective method, although this may come with additional caveats resulting from the associated reduction in sensitivity (when cases are not detected) and specificity (when students self-isolate but are not infected) [45]. Additionally, in higher educational settings, it is important to consider any impact on teaching and examination schedules as well as mental health and well-being of students. The models considered here allow for an estimate of the different resources used by the different control strategies. In order to determine an optimal intervention, it is crucial to establish the objective of any control policy, noting that the objective may not be generalizable across all higher education establishments. Once an objective is appropriately defined, any modelling can be specifically tailored to maximize the robustness of any advice offered.

Furthermore, a growing picture is just beginning to emerge on the prevalence of, and risk factors for, 'long COVID' symptoms and health complications following coronavirus (COVID-19) infection. An initial set of early experimental results collected by the ONS indicates around one in five respondents testing positive for COVID-19 exhibit symptoms for a period of five weeks or longer, and around one in 10 respondents testing positive for COVID-19 exhibit symptoms for a period of 12 weeks or more [46,47]. We recognize that the current university closures may have significant impact upon student mental health and well-being—across multiple surveys collecting information on how the COVID-19 pandemic has affected the mental health of students, a consistent outcome was above 50% of respondents expressing that their well-being and mental health had become worse [48]. In addition, we hope that the ongoing vaccine roll-out will provide a level of protection for those most vulnerable to severe outcomes, which in turn may alleviate risks associated with possible student to community spread.

In conclusion, our findings are comprised of three overarching points. Firstly, we observed evidence of spillover transmission between higher education populations and the wider community in some, but not all, settings. Secondly, we would expect reductions in adherence to NPIs (including case and household isolation) to have more impact than any marginal benefits generated from a staggered return of students to university. Thirdly, the emergence of more transmissible new variants results in impaired effectiveness of mass asymptomatic testing. Ultimately, we hope that the work presented here can be used by universities and policymakers to assist in the long-term strategy of ensuring that students can return safely to their studies at universities in the UK. And while we have focused on the national picture in the UK, we also hope our results can offer insights relevant to higher education in other countries.

Data accessibility. This article has no additional data.

Authors' contributions. J.E. conceptualization, methodology, software, formal analysis, investigation, writing—original draft, review and editing, visualization. E.M.H. conceptualization, methodology, software, formal analysis, investigation, writing—original draft, review and editing, visualization. H.B.S. conceptualization, methodology, software, formal analysis, investigation, writing—original draft, review and editing, visualization. K.J.B. conceptualization, methodology, investigation, data curation, writing—original draft, review and editing, visualization, supervision. E.J.N. conceptualization, methodology, software, formal analysis, investigation, writing—original draft, review and editing, visualization. E.L.F. conceptualization, methodology, software, formal analysis, investigation, data curation, writing—original draft, review and editing, visualization. M.L.T. conceptualization, investigation, writing—original draft, review and editing. E.B.-P. conceptualization, software, review and editing, supervision. L.D. conceptualization, review and editing, supervision. C.J.B. conceptualization, methodology, software, formal analysis, investigation, writing—original draft, review and editing, visualization. R.B.H. conceptualization, methodology, software, formal analysis, investigation, writing—original draft, review and editing, visualization. L.S. conceptualization, methodology, software, formal analysis, investigation, writing—original draft, review and editing, visualization. J.R.G. conceptualization, writing—original draft, review and editing, supervision, project management. M.J.T. conceptualization, methodology, formal analysis, investigation, writing—original draft, review and editing, supervision, project management.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by EPSRC grant no EP/R014604/1. The authors would also like to thank the Virtual Forum for Knowledge Exchange in Mathematical Sciences (V-KEMS) for the support during the workshop *Unlocking higher education Spaces – What Might Mathematics Tell Us?* where work on this paper was undertaken. K.J.B.

acknowledges support from a University of Nottingham Anne McLaren Fellowship. E.L.F. acknowledges support via K.J.B.'s fellowship and the Nottingham BBSRC Doctoral Training Partnership. M.L.T. was supported by the UK Engineering and Physical Sciences Research Council (grant no. EP/N509620/1). E.B.-P., E.J.N., L.D., J.R.G. and M.J.T. were supported by UKRI through the JUNIPER modelling consortium (grant no. MR/V038613/1). E.M.H., L.D. and M.J.T. were supported by the Medical Research Council through the COVID-19 Rapid Response Rolling Call (grant no. MR/V009761/1). H.B.S. is funded by the Wellcome Trust and the Royal Society (grant no. 202562/Z/16/Z). J.E. is partially funded by the UK Engineering and Physical Sciences Research Council (grant no. EP/T004878/1).

Acknowledgements. The authors thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support during the programme *Infectious Dynamics of Pandemics* where work on this paper was undertaken.

Appendix A. A simple outbreak model for university COVID-19 outbreaks

We postulate in §2.1 that the probability, $\mathcal{P}$, that a university experiences a SARS-CoV-2 outbreak is given by

$$\mathcal{P} = 1 - p^n, \tag{A 1}$$

where n is the number of imported cases and p is the probability that an imported case fails to seed an outbreak.

We tested this hypothesis by using estimates for the number of imported student cases [7] and COVID-19 case number data for a number of universities for which cumulative case number data was available on the UCU COVID dashboard [8].

For a university i with cumulative case number c_i, we defined an outbreak if $c_i > T_u$, where T_u is a threshold number of cases. We set $x_i = 1$ if a university had experienced an outbreak, and $x_i = 0$ if not.

The probability mass function for the distribution of outbreaks is given by

$$f(x|p) = p^{n(1-x)}(1 - p^n)^x.$$

Thus, the likelihood of the data for all N universities, given p, is

$$L(p) = \prod_{i=1}^{N} p^{n_i(1-x_i)}(1 - p^{n_i})^{x_i}$$

and the log likelihood is

$$LL(p) = \sum_{i=1}^{N} \{n_i(1 - x_i)\log p + x_i \log(1 - p^{n_i})\}.$$

Maximizing the log likelihood gives the maximum-likelihood estimate $\hat{p}$ for p. The $100(1 - \alpha)\%$ confidence interval for p is given by

$$\left[\hat{p} \pm z(\alpha/2)\frac{1}{\sqrt{NI(\hat{p})}}\right],$$

where $z(\alpha/2)$ is defined by $P(Z > z(\alpha/2)) = \alpha/2$ for $Z \sim N(0, 1)$ and

$$I(p) = -E\left(\frac{\mathrm{d}^2}{\mathrm{d}\,p^2}LL(p)\right) \tag{A 2}$$

$$= E\left(\sum_{i=1}^{N}\frac{n_i}{p^2(1 - p^{n_i})^2}\{(1 - p^{n_i})^2 - x_i(1 - p^{n_i}) + x_i n_i p^{n_i}\}\right) \tag{A 3}$$

$$= \sum_{i=1}^{N}\frac{n_i^2 p^{n_i}}{p^2(1 - p^{n_i})}, \tag{A 4}$$

using $E(x_i) = 1 - p^{n_i}$.

Table 7. Coefficients, and associated *p*-value and standard error, for the univariate and intermediate multivariate logistic regression models for hall SAR.

covariate	coefficient	*p*-value	s.e.
univariate logistic regression: SAR			
hall size	0.0037	<0.0001	0.00006
constant	−2.8388	0.0001	0.1722
median household size	−0.0539	0.0029	0.0181
constant	−1.3218	<0.0001	0.01590
proportion shared bathroom	0.3541	0.0017	0.1097
constant	−1.9836	<0.0001	0.0822
proportion medical faculty	0.3511	0.0004	1.7973
constant	−2.5257	<0.0001	0.2238
multivariate logistic regression: SAR			
hall size	0.0030	<0.0001	0.0007
median household size	0.0300	0.2458	0.0258
proportion shared bathroom	0.4141	0.0010	0.1253
proportion medical faculty	4.0628	0.0712	2.2521
constant	−3.6647	<0.0001	0.4690
hall size	0.0030	<0.0001	0.0007
proportion shared bathroom	0.3977	0.0013	0.1233
proportion medical faculty	2.7342	0.1588	1.9402
constant	−3.2354	<0.0001	0.2795

Appendix B. Additional analyses for student hall infection data

B.1. Additional regression results

Univariate and intermediate multivariate regression results for the household and hall SAR (§2.2) are summarized in tables 7 and 8.

B.2. Stochastic transmission model for hall and household infection

An alternative method for exploring the role of household and hall size, discussed briefly in §2.2, is to fit a stochastic transmission model that allows for infection between hall members in addition to household members.

B.2.1. Methods

We first calculated the household size distribution for each hall. We ignore the temporal dynamics (setting the infectious period to unity) and simulated the final size of the outbreak using the Sellke construction [49] in a population with two levels of mixing, defined by the household infectious contact rate, λ_H, and the global (or hall) infectious contact rate, λ_G. Motivated by the lack of dependence of household SAR on household size (table 1), we assumed density-dependent mixing in households. Contacts at each level were assumed to be made at the points of a homogeneous Poisson process. We calculated the probability of a student being infected given a single introduction in the hall. Inference was performed for each hall using the approximate Bayesian computation tutorial in Kypraios *et al.* [50], assuming Exp(1) priors for λ_H and λ_G.

B.2.2. Results

In figure 18*a*, we plot the probability that a student was infected by another within their hall including their household ($p_{\text{hall}} = 1 - e^{-\lambda_G AR}$) (where AR is the hall attack rate, in this case the number of reported

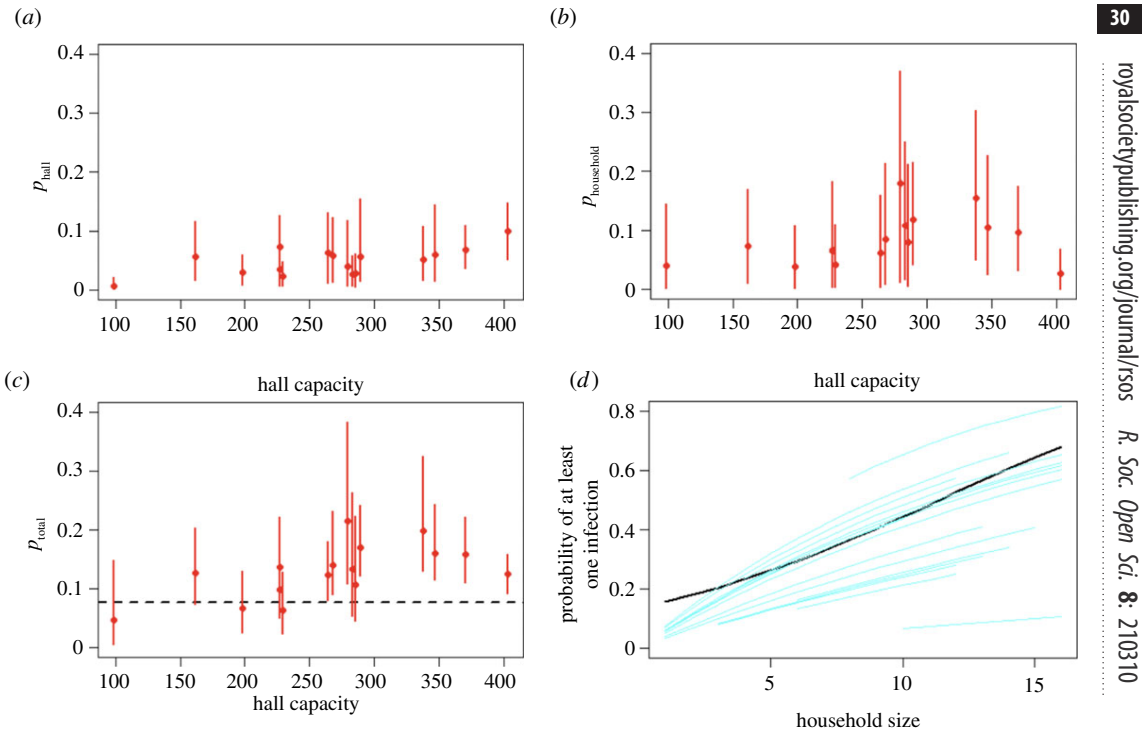

Figure 18. Results of fitting the model with two levels of mixing to each hall individually, plotted against hall capacity. Circles indicate expected mean and lines 95% confidence intervals. (*a*) Probability of infection due to global infectious contact. (*b*) Probability of infection due to household infectious contact. (*c*) Comparison of the total probability of infection after introduction accounting for household and global infectious contacts compared with the estimated binomial probability of infection given introduction into a household (black dashed line). (*d*) Comparison of the probability of infection in a household by household size for each hall (blue lines) and the output from the binary regression analysis (black line).

Table 8. Coefficients, and associated *p*-value and standard error, for the univariate and intermediate multivariate logistic regression models for household SAR.

covariate	coefficient	*p*-value	s.e.
univariate logistic regression: SAR			
household size	−0.0543	0.1442	0.0372
constant	−2.2885	<0.0001	0.3679
date of first infection	−0.1354	<0.0001	0.0291
constant	−0.8816	0.0272	0.3996
proportion shared bathroom	0.7472	0.0151	0.3074
constant	−3.3660	<0.0001	0.2730
multivariate logistic regression: SAR			
household size	−0.0743	0.0558	0.0388
date of first infection	−0.1547	<0.0001	0.0305
proportion shared bathroom	0.9911	0.0015	0.3115
constant	−0.6350	0.2966	0.6084

confirmed infections with known household). The additional probability of infection from within their household is shown as $p_{\text{household}} = 1 - e^{-\lambda_H}$ in figure 18*b*. In figure 18*c*, we plot the probability of infection (p_{total}) accounting for both household and global infectious contacts within an infected household. This is compared with the binomial probability of reporting an infection given a

previously reported infection in a household (which does not distinguish between halls). Figure 18*d* compares the estimated probability of a household reporting an infection for each hall to the estimation from the binomial logistic analysis (table 3). There is some indication that global infectious contacts may play a relatively greater role in overall infection risk in the largest halls. However, choices for distributing missing household data, which is ignored for here, will probably influence the relative size of p_{hall} and $p_{\text{household}}$, as will choices about scaling of mixing intensity with household size.

The maximum hall size in this data is approximately 400 students and findings may not generalize to other hall settings or future periods of student return. Other limitations of this approach are the lack of differentiation between symptomatic and asymptomatic infections, pre-existing immunity, or the impact of isolation, so that parameters are interpreted as averages across students in a hall in addition to the caveats arising from the missing data. Furthermore, we assume a single introduction and a closed system of fixed occupancy, so that any imported cases are attributed to infection within the hall. Dedicated household-based studies in student residential halls would be valuable for untangling the role of mixing within households, halls and with the community on infection risk in these settings.

Appendix C. Additional information on age-stratified observations

C.1. Additional observations: age-stratified analysis

C.1.1. Methodological details

The numerical interpolation method, and the subsequent calculation of the growth rate of positive cases, is applied to the positive case counts in each LTLA, rescaled by the number of people (falling within the considered age range) estimated to live there. This quantity, $c(t)$, shows a consistent day-of-week effect due to e.g. varying test availability and test seeking behaviour. To account for overdispersion in the data, we assume a quasi-Poisson distribution in the fitting.

A smoother $\varrho(t)$ is applied using thin-plate splines, such that $c(t) \propto e^{\varrho(t)+\omega_i}$, where $\omega_i \; \forall \; i \in [1, 7]$ is used to apply a fixed effect for each day of the week. The instantaneous growth rate of the cases is simply given by $\varrho'(t)$. This was implemented using a general additive model from the R package *mgcv* with a canonical link [51]. Past examples of this method can be found in [52] (see figure 19).

C.1.2. Growth rates

Despite the clear spikes in cases among 18–24-year-olds across all LTLAs in figure 20, the growth rates for community cases are qualitatively very different. In figure 20*c*, the community growth rate mirrors the national trend. In figure 20*a*, the growth rate of community cases is higher, and appears to lag after the growth in student-aged cases.

In figure 20*d*, a qualitatively different scenario emerges, with a marked rise in the growth of community cases following an outbreak among the student-aged population. Finally, in figure 20*b*, the outbreak among 18–24-year-olds has no perceptible impact on the growth rate of community cases.

C.1.3. Limitations

The estimated growth rates of confirmed cases, and the estimated excess community cases following a large student-aged outbreak, are sensitive to the choice of the spline in the smoother. Changing the spline does not qualitatively alter our conclusions.

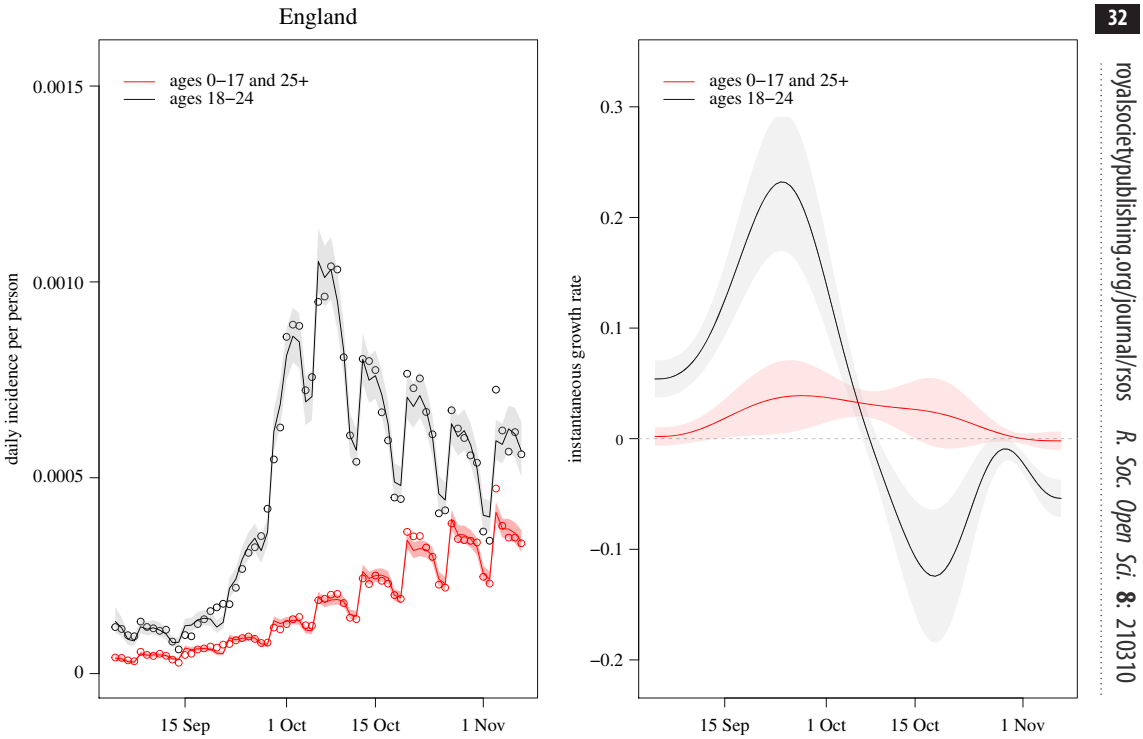

Figure 19. Growth rate among the student-aged (grey) and community populations (red) across England. University outbreaks are observed on the national scale, with the higher incidence per population among those aged 18–24. The community growth rate in cases increased from late September 2020. However, there was not a statistically significant subsequent increase following the peak in student-aged outbreaks. The shaded regions are the 95% confidence intervals for the relevant quantity.

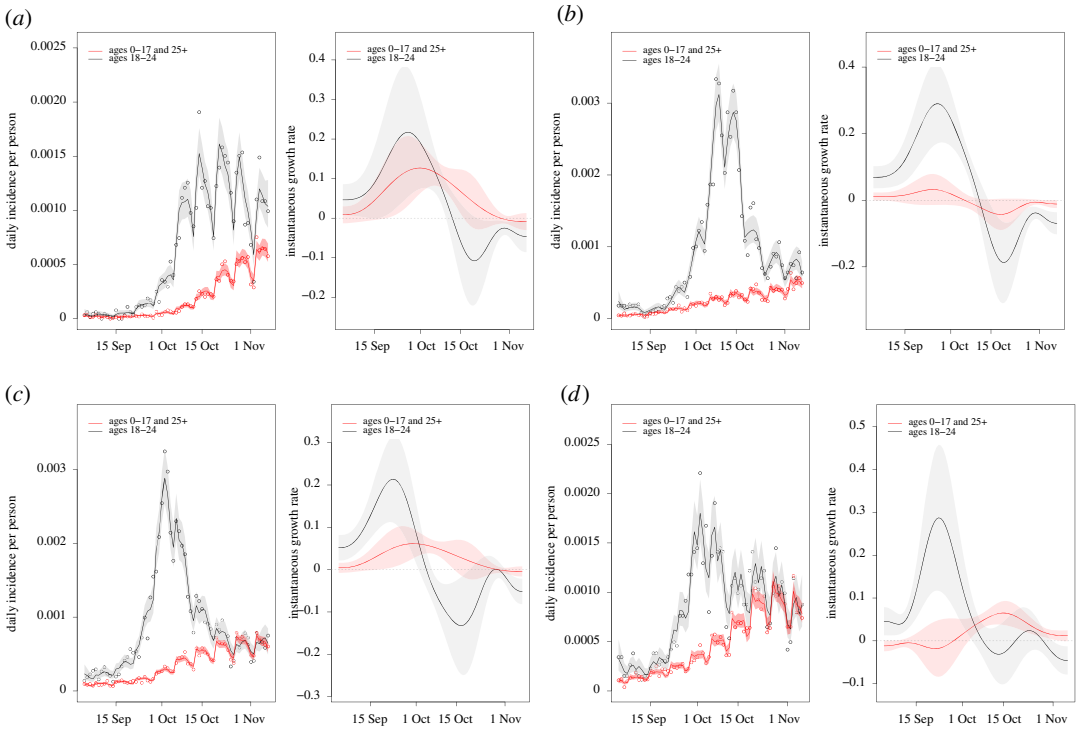

Figure 20. Examples of the different types of growth rate patterns observed among student-aged (grey) and community cases (red). The shaded regions are the 95% confidence intervals for the relevant quantity. (*a*) Bristol, (*b*) Durham, (*c*) Leeds and (*d*) Salford.

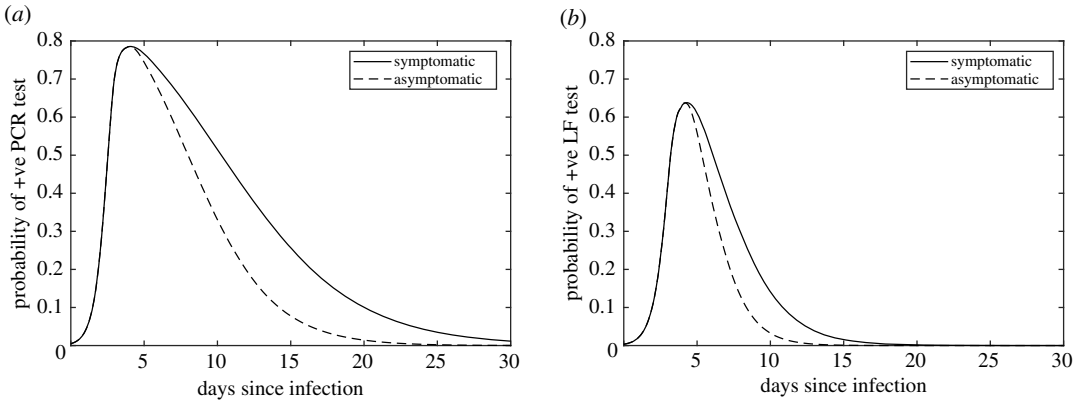

Figure 21. Probabilities of testing positive through time for symptomatic and asymptomatic individuals. We assumed that the probability of positive test results being returned in symptomatic and asymptomatic individuals were equal during the proliferation stage of the virus, but that the probability of asymptomatic individuals testing positive decayed faster in the clearance stage, owing to a shorter mean clearance duration of 6.7 days [53] (*a*) PCR test; (*b*) LFT.

Appendix D. Additional information on the network-based structured model

D.1. Test sensitivity

The probability of testing positive is probably a function of viral load; while symptomatic and asymptomatic individuals have similar average peak viral loads and proliferation stage durations, their average duration of clearance stages has been observed to differ [53,54]. Therefore, we used distinct test sensitivity profiles for symptomatic and asymptomatic cases (figure 21). However, we highlight that this is an area of considerable uncertainty. Future studies detailing the testing probability of asymptomatic individuals, and the specific relationship between viral load and testing probability, would be a valuable contribution to this area.

For symptomatic cases, we used posterior median profiles reported by Hellewell *et al.* [55] of the probability of detecting infection against time since infection, with separate estimates for PCR and lateral flow tests (LFTs). The analysis used cycle threshold (Ct) data from repeat PCR testing of healthcare workers in the SAFER study [56], with infections confirmed by paired serology. The probability of detection by LFT was estimated given an assumption that an LFT would detect infections with a $Ct \leq 27$.

For asymptomatic cases, we assumed that the probability of asymptomatic individuals testing positive is equal to that of symptomatic individuals until the peak of infection, but then decays more rapidly, such that the probability of an asymptomatic individual testing positive at 6.7 days after the peak should equal the probability of a symptomatic individual testing positive at 10.5 days after the peak (corresponding with findings from Kissler *et al.* [53] who estimated an average duration of clearance of 10.5 days in symptomatic cases versus 6.7 days in asymptomatic cases) (figure 21).

The sensitivity of PCR tests when conditioned on having received a positive LFT result may differ from the sensitivity estimates of an independent PCR test. We assumed that individuals receiving a positive LFT result would be certain to return a positive result from the confirmatory PCR test.

D.2. Test specificity

We assumed the specificity of PCR tests to be 100%, in line with the ONS UK COVID-19 Infection Survey indicating the specificity of the used PCR tests being in excess of 99.9% [57,58]). We assumed LFT specificity to be 99.68% [59]. Using LFTs to test entire year groups, false positives would be expected to occur relatively frequently.

D.3. Model change log

We detail here notable parameter changes and additions to the previously presented network model [27].

D.3.1. Isolation length

From 14 December 2020, the guidance from the UK government on the period of isolation for contacts of confirmed cases was reduced from 14 days to 10 days. The corresponding periods of isolation have been revised in the model.

D.3.2. Infection risk for students awaiting return to university

For susceptibles not yet returned to the university, we computed a daily probability of infection to give a background prevalence of between 0.5% and 2% (with an infection duration of 16 days, across latent and infectious periods). We sampled the background prevalence in each simulation replicate from a Uniform (0.005, 0.02) distribution.

D.3.3. Proportion of individuals who stayed in university accommodation between terms

Student surveys indicated that of the order of 10% of students intended to stay in their university accommodation after the end of the first academic term [48].

In each simulation replicate, we sampled the proportion independently for on-campus and off-campus residents from a Uniform (0.05,0.15) distribution, thus ensuring we included uncertainty associated with this quantity across our collection of simulations.

D.3.4. Contact patterns

We applied the following two contact pattern changes to all but the baseline (no intervention) scenario: (i) society contacts did not occur (transmission risk therefore zero), with it assumed that all meetings would take place online; (ii) for on-campus resident students, we set a zero probability of a contact being made with an individual within the broader accommodation unit of the same floor or block of residence (thus outside the immediate household).

D.3.5. Fraction of previous infecteds with PCR positive test result in the previous 90 days

In each simulation replicate, we sampled the fraction of previous infecteds who had returned a PCR positive test result in the previous 90 days from a uniform distribution, Uniform(0.02,0.05).

Individuals set as being present in accommodation prior to the start of the simulation entered the return testing procedure in an equivalent way to individuals with later arrival dates, with entry time determined by the relevant staggered student return strategy. For individuals from this group that became symptomatic and received a positive test result in the gap before their envisaged time to begin the return test process, they satisfied the condition of having had a positive PCR result within the previous 90 days and, as a consequence, no longer underwent the return test process.

D.3.6. Assumptions for scenarios related to isolation status under staggered return and leaving return testing process

Returning students that have symptoms are by definition non-adherent to guidance. In this situation, for the household the returning student is joining, other adhering household members may enter household isolation. We assumed any such individuals entered isolation for the full 10-day period, irrespective of the date of symptom onset of the symptomatic individual.

In the scenario of a student completing the return testing procedure with negative results, and who would be entering a household that had household members in isolation due to the presence of a recently confirmed case, the student leaving the return test process would immediately enter household isolation.

Appendix E. Additional information on the asymptomatic screening model

For this analysis, we used a layered network model of contacts between 15 000 simulated students, with one layer of household contacts and one of other-group contacts intended to simulate all out-of-household contact. We start the simulation with 100 infectious individuals, and run the model for 100 timesteps (notionally days). For each scenario, we plot the results of 100 replicates, each run on a newly generated network. Importantly: the particular parameters for this model have been chosen for a combination of plausibility and simplicity, and some are not well-founded in any particular dataset (we attempt to highlight these).

Half of the households were of 10 people, and half of 5 people (to simulate a cluster-flat arrangement in large halls, e.g. [60]). Other-group contacts are added in 3000 groups, with 5% of groups of size 40, 30% of size 10, 50% of size 5 and 15% of size 3—these values were chosen to simulate a range of activities, but are not well-founded in data. Results are not sensitive to small perturbations in these group sizes, but are sensitive to large changes in the overall amount of group contact. Within either household or other groups all individuals are assumed to have pairwise contact at all timesteps when the individuals are not isolating.

Disease progression and isolation are governed by a stochastic rate-based compartmental model in which individuals can be susceptible, exposed but not yet infectious, presymptomatically infectious, asymptotically infectious, symptomatically infectious, or recovered (and presumed immune). They can also be in these various states and self-isolating with their household. Individuals become exposed when one of their network contacts infects them—here household contacts have a 2.5% day^{-1} probability of infecting each of their susceptible household members (note that this is independent of household size), and non-household contacts transmit with 1/10th this probability. These probabilities are increased by a factor of 1.5 when simulating a more-transmissible variant. These transmission figures have been chosen for simplicity and to plausibly reflect reasonable within-household attack rates. Where no other citation is given, rates of progression between disease states are round-number versions of the fitted parameters from [61]. Exposed individuals become presymptomatically or asymptomatically infectious at a rate of $0.33\,\text{day}^{-1}$ to give a mean 3-day latent period. Presymptomatically infectious individuals become symptomatic at a rate of $0.5\,\text{day}^{-1}$ to give a mean 2-day presymptomatic period. Symptomatically infectious people recover at a rate of $0.1\,\text{day}^{-1}$ to give a mean symptomatic infectious period of 10 days, a round-number version of the 9.5 days reported in [62]. We do not include hospitalization or death, as these events are very rare in the young-adult population. Half of infected individuals are assumed to develop symptoms, and half to remain asymptomatic (or non-test-seeking for some other reason). Asymptomatic individuals are infectious for the same mean total period of time as symptomatic individuals, and are equally infectious—predictably the effectiveness of asymptomatic screening is sensitive to this assumption.

Both symptomatic and asymptomatic testing are assumed to be perfect and rapid, returning results on the day of testing and giving neither false positives nor false negatives. Symptomatic individuals are assumed to immediately seek testing on the day symptoms develop. When an individual receives a positive test, they and their entire household are assumed to isolate perfectly from all non-household contacts, but continue to interact with household contacts as before. Non-household contacts of test-positives are traced and isolated with probability 0.5.

This model is an adaptation of a model originally written to model COVID-19 in Caribbean communities, available at https://github.com/SaraJakubiak/covid19-caribbean-educational-model—the majority of features within that model (including dynamically changing network, age-structure, etc.) are not used here. The adaptation of this code to the HE setting used to produce these results can be found at: https://github.com/magicicada/covid19-caribbean-educational-model/tree/manuscript-INI-HE-group.

References

1. Higher Education Statistics Agency. Who's studying in HE? 2020. See https://www.hesa.ac.uk/data-and-analysis/students/whos-in-he (visited on 4 February 2021).
2. Virtual Forum for Knowledge Exchange in the Mathematical Sciences (V-KEMS). *Unlocking Higher Education Spaces – What Might Mathematics Tell Us?* 2020. See https://gateway.newton.ac.uk/sites/default/files/asset/doc/2007/Unlocking%20HE%20Spaces%20July%202020_small_0.pdf (visited on 4 February 2021).
3. Isaac Newton Institute for Mathematical Sciences. Infectious Dynamics of Pandemics: Mathematical and statistical challenges in understanding the dynamics of infectious disease pandemics. 2020. See https://www.newton.ac.uk/event/idp (visited on 4 February 2021).
4. Department for Education. Press release: Updated guidance for universities ahead of reopening. 2020. See https://www.gov.uk/government/news/updated-guidance-for-universities-ahead-of-reopening (visited on 3 February 2021).
5. Task and Finish Group on Higher Education/Further Education. Principles for managing SARS-CoV-2 transmission associated with higher education, 3 September 2020. 2020. See https://www.gov.uk/government/publications/principles-for-managing-sars-cov-2-transmission-associated-with-higher-education-3-september-2020 (visited on 5 February 2021)).
6. Brooks-Pollock E *et al.* 2020 High COVID-19 transmission potential associated with re-opening universities can be mitigated with layered interventions. *medRxiv*. (doi:10.1101/2020.09.10.20189696).
7. Enright J. 2020 Basic arriving student calculations. See https://github.com/magicicada/simple_epi_calculations/blob/master/basic_arriving_student_calculations/output_estimate_tables/sept25_estimate_number_incoming_infected_from_regions_combined.csv (visited on 11 November 2020).
8. UCU. Monitoring cases of Covid-19 in UK higher and further education. 2020. See https://www.ucu.org.uk/covid-dashboard (visited on 19 November 2020).
9. Endo A, Abbott S, Kucharski AJ, Funk S. 2020 Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China.

Wellcome Open Res. **5**, 67. (doi:10.12688/wellcomeopenres.15842.3)

10. The MathWorks, Inc. *MATLAB and Statistics Toolbox Release R2018*. 2018. See https://uk.mathworks.com/products/matlab.html.
11. VSN International Ltd. *Genstat for Windows 21st edn*. 2020. See https://vsni.co.uk/software/genstat.
12. Lyngse FP, Kirkeby CT, Halasa T, Andreasen V, Skov RL, Møller FT, Krause TG, Mølbak K. 2020 COVID-19 transmission within Danish households: a nationwide study from lockdown to reopening. *medRxiv*. (doi:10.1101/2020.09.09.20191239)
13. Toth DJ, Beams AB, Keegan LT, Zhang Y, Greene T, Orleans B, Seegert N, Looney A, Alder SC, Samore MH. 2021 High variability in transmission of SARS-CoV-2 within households and implications for control. *medRxiv*. (doi:10.1101/2021.01.29.20248797)
14. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. 2020 Household Transmission of SARS-CoV-2: a systematic review and meta-analysis. *JAMA Netw. Open* **3**, e2031756–e2031756. (doi:10.1001/jamanetworkopen.2020.31756)
15. Office for National Statistics. Population estimates by output areas, electoral, health and other geographies, England and Wales: mid-2019. 2020. See https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualsmallareapopulationestimates/mid2019 (visited on 5 February 2021).
16. Office for National Statistics. COVID-19 Infection Survey. 2020. See https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata/2020 (visited on 8 February 2021).
17. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D, Fraser C. 2020 Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936. (doi:10.1126/science.abb6936)
18. Office for National Statistics. Economic activity of full-time students by student accommodation by age. 2020. See https://www.nomisweb.co.uk/census/2011/dc6108ew (visited on 4 February 2021).
19. Public Health England. GOV.UK Coronavirus (COVID-19) in the UK. 2020. See https://coronavirus.data.gov.uk/details/download (visited on 4 February 2021).
20. Office for National Statistics. Middle Layer Super Output Areas (December 2011) Boundaries Full Clipped (BFC) EW V3. 2021. See https://geoportal.statistics.gov.uk/datasets/middle-layer-super-output-areas-december-2011-boundaries-full-clipped-bfc-ew-v3 (visited on 5 February 2021).
21. Department for Education. Press release: All students offered testing on return to university. 2020. See https://www.gov.uk/government/news/all-students-offered-testing-on-return-to-university (visited on 3 February 2021).
22. Department for Education. Students returning to, and starting higher education, in spring term 2021. 2020. See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957926/HE_guidance_spring_term_guidance.pdf (visited on 3 February 2021).
23. The COVID-19 Genomics UK (COG-UK) consortium. Lineage-specific growth of SARS-CoV-2 B.1.1.7 during the English national lockdown. 2021. See https://virological.org/t/lineage-specific-growth-of-sars-cov-2-b-1-1-7-during-the-english-national-lockdown/575 (visited on 3 February 2021).
24. Public Health England. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01 Technical briefing 3. 2021. See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950823/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3_-_England.pdf (visited on 3 February 2021).
25. Davies NG *et al.* 2020 Estimated transmissibility and severity of novel SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, 6538. (doi:10.1126/science.abg3055)
26. Volz E *et al.* 2021 Transmission of SARS-CoV-2 lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. *medRxiv*. (doi:10.1101/2020.12.30.20249034).
27. Hill EM, Atkins BD, Keeling MJ, Tildesley MJ, Dyson L. 2021 Modelling SARS-CoV-2 transmission in a UK university setting. *Epidemics* **36**, 100476. (doi:10.1016/j.epidem.2021.100476)
28. Danon L, House TA, Read JM, Keeling MJ. 2012 Social encounter networks: collective properties and disease transmission. *J. R. Soc. Interface* **9**, 2826–2833. (doi:10.1098/rsif.2012.0357)
29. Danon L, Read JM, House TA, Vernon MC, Keeling MJ. 2013 Social encounter networks: characterizing Great Britain. *Proc. R. Soc. B* **280**, 20131037. (doi:10.1098/rspb.2013.1037)
30. University of Bristol. Data protection policy. 2021. See http://www.bristol.ac.uk/media-library/sites/secretary/documents/information-governance/data-protection-policy.pdf (visited on 2 February 2021).
31. Keeling MJ, Dyson L, Guyver-Fletcher G, Holmes A, Semple MG, Tildesley MJ, Hill EM. 2020 Fitting to the UK COVID-19 outbreak, short-term forecasts and estimating the reproductive number. *medRxiv*. (doi:10.1101/2020.08.04.20163782).
32. Wikipedia. Violin Plot. 2021. See https://en.wikipedia.org/wiki/Violin_plot.
33. University of Bristol. Courses with in-person activity in Teaching Block 2 until 26 March. 2021. See http://www.bristol.ac.uk/students/your-studies/study-2020/your-course/teaching-weeks/ (visited on 12 January 2021).
34. University of Warwick. Term 2 update. 2020. Available from: https://warwick.ac.uk/insite/news/intnews2/student-return/ (visited on 27 January 2021).
35. Hambridge HL, Kahn R, Onnela JP. 2021 Examining SARS-CoV-2 interventions in residential colleges using an empirical network. *medRxiv*. (doi:10.1101/2021.03.09.21253198)
36. Eames KT, Tang ML, Hill EM, Tildesley MM, Read JM, Keeling MJ, Gog JR. 2021 Coughs, colds and 'freshers' flu' survey in the University of Cambridge, 2007–2008. *medRxiv*. (doi:10.1101/2021.03.31.21251220)
37. Office for National Statistics. Coronavirus and the impact on students in higher education in England: September to December 2020. 2020. See https://www.ons.gov.uk/peoplepopulationandcommunity/educationandchildcare/articles/coronavirusandtheimpactonstudentsinhighereducationinenglandseptembertodecember2020/2020-12-21#student-behaviour.
38. Nixon E *et al.* 2021 Contacts and behaviours of university students during the COVID-19 pandemic at the start of the 2020/2021 academic year. *Sci. Rep.* **11**, 1–13. (doi:10.1038/s41598-021-91156-9)
39. Lopman B, Liu CY, Le Guillou A, Handel A, Lash TL, Isakov AP, Jenness SM. 2021 A modeling study to inform screening and testing interventions for the control of SARS-CoV-2 on university campuses. *Sci. Rep.* **11**, 1–11. (doi:10.1038/s41598-021-85252-z)
40. Goyal R, Hotchkiss J, Schooley RT, De Gruttola V, Martin NK. 2021 Evaluation of severe acute respiratory syndrome Coronavirus 2 transmission mitigation strategies on a university campus using an agent-based network model. *Clin. Infect. Dis.* (doi:10.1093/cid/ciab037)
41. Harper PR, Moore JW, Woolley TE. 2021 Covid-19 transmission modelling of students returning home from university. *Health Syst.* **10**, 31–40. (doi:10.1080/20476965.2020.1857214)
42. de Figueiredo A. 2021 Forecasting sub-national trends in COVID-19 vaccine uptake in the UK. *medRxiv*. (doi:10.1101/2020.12.17.20248382)
43. Sandmann FG *et al.* 2021 The potential health and economic value of SARS-CoV-2 vaccination alongside physical distancing in the UK: a transmission model-based future scenario analysis and economic evaluation. *Lancet Infect. Dis.* **21**, 962–974. (doi:10.1016/S1473-3099(21)00079-7)
44. Tildesley MJ *et al.* 2021 Optimal health and economic impact of non-pharmaceutical intervention measures prior and post vaccination in England: a mathematical modelling study. *medRxiv* (doi:10.1101/2021.04.22.21255949).
45. Hemani G *et al.* 2021 Modelling pooling strategies for SARS-CoV-2 testing in a university setting. *Wellcome Open Res.* **6**, 70. (doi:10.12688/wellcomeopenres.16639.1)
46. Office for National Statistics. The prevalence of long COVID symptoms and COVID-19 complication. 2020. See https://www.ons.gov.uk/news/statementsandletters/theprevalenceoflongcovidsymptomsandcovid19complications (visited on 5 February 2021).
47. Office for National Statistics. Updated estimates of the prevalence of long COVID symptoms. 2021. See https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/adhocs/12788updatedestimatesoftheprevalenceoflongcovidsymptoms (visited on 5 February 2021).
48. Office for National Statistics. Coronavirus and the impact on students in higher education in

England: September to December 2020. 2020. See https://www.ons.gov.uk/peoplepopulationandcommunity/educationandchildcare/articles/coronavirusandtheimpactonstudentsinhighereducationinenglandseptembertodecember2020/2020-12-21 (visited on 5 February 2021).

49. Diekmann O, Heesterbeek JAP. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Wiley Series in Mathematical & Computational Biology. Wiley.
50. Kypraios T, Neal P, Prangle D. 2017 A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. *Math. Biosci.* **287**, 42–53. (doi:10.1016/j.mbs.2016.07.001)
51. Wood SN. 2019 Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. https://cran.r-project.org/web/packages/mgcv/ (doi:10.1201/9781315370279).
52. Pellis L *et al.* 2020 Challenges in control of Covid-19: short doubling time and long delay to effect of interventions. *Phil. Trans. R. Soc. B* **376**, 20200264. (doi:10.1098/rstb.2020.0264)
53. Kissler SM *et al.* 2020 SARS-CoV-2 viral dynamics in acute infections. *PLoS Biol.* **19**, e3001333. (doi:10.1371/journal.pbio.3001333)
54. Uhm JS *et al.* 2020 Patterns of viral clearance in the natural course of asymptomatic COVID-19: Comparison with symptomatic non-severe COVID-19. *Int. J. Infect. Dis.* **99**, 279–285. (doi:10.1016/j.ijid.2020.07.070)
55. Hellewell J *et al.* 2020 Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *BMC Med.* **19**, 106. (doi:10.1186/s12916-021-01982-x)
56. Houlihan CF *et al.* 2020 Pandemic peak SARS-CoV-2 infection and seroconversion rates in London frontline health-care workers. *Lancet* **396**, e6–e7. (doi:10.1016/S0140-6736(20)31484-7)
57. Office for National Statistics. COVID-19 Infection Survey (Pilot): methods and further information. 2020. See https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation#test-sensitivity-and-specificity (visited on 5 February 2021).
58. Office for National Statistics. Coronavirus (COVID-19) Infection Survey, UK: 18 December 2020. 2020. See https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/18december2020#test-sensitivity-and-specificity (visited on 5 February 2021).
59. Joint PHE Porton Down & University of Oxford SARS-CoV-2 test development and validation cell. Preliminary report from the Joint PHE Porton Down & University of Oxford SARS-CoV-2 test development and validation cell: Rapid evaluation of Lateral Flow Viral Antigen detection devices (LFDs) for mass community testing; 2020. See https://www.ox.ac.uk/sites/files/oxford/media_wysiwyg/UK%20evaluation_PHE%20Porton%20Down%20%20University%20of%20Oxford_final.pdf (visited on 5 February 2021).
60. University of Glasgow. Murano Street Student Village; 2021. See https://www.gla.ac.uk/undergraduate/accommodation/residenceprofiles/muranostreetstudentvillage/ (visited on 4 February 2021).
61. Banks CJ *et al.* 2020 Disentangling the roles of human mobility and deprivation on the transmission dynamics of COVID-19 using a spatially explicit simulation model. *medRxiv* (doi:10.1101/2020.11.25.20144139).
62. Hu Z *et al.* 2020 Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci. China Life Sci.* **63**, 706–711. (doi:10.1007/s11427-020-1661-4)

CHAPTER 6

Influence of setting-dependent contacts and protective behaviours on asymptomatic SARS-CoV-2 infection amongst members of a UK university

Note: This chapter is under review as a research article.

Supplementary material for this chapter can be found in Appendix G.

Author contributions: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Influence of setting-dependent contacts and protective behaviours on asymptomatic SARS-CoV-2 infection amongst members of a UK university

Emma Fairbanks[1,2], Kirsty Bolton[2,*], Ru Jia[3], Grazziela Figueredo[4], Holly Knight[3], and Kavita Vedhara[3]

[1]School of Veterinary Medicine and Science, University of Nottingham
[2]School of Mathematical Sciences, University of Nottingham
[3]School of Medicine, University of Nottingham
[4]School of Computer Science, University of Nottingham
[*]Corresponding author: kirsty.bolton@nottingham.ac.uk

Abstract

Background: The risk of SARS-CoV-2 infection during university activities is poorly understood.

Methods: We survey 62 users of a university asymptomatic SARS-CoV-2 testing service on details of their activities and contacts in the 7 days prior to receiving a positive or negative SARS-CoV-2 PCR test result. After classifying activities into settings, Bayesian logistic regression is used to model test outcome based on a set of setting-specific contact measures for a variety of contact definitions. For each contact measure, support intervals based on Bayes factors are computed to identify settings for which a contact definition is associated with test outcome. Posterior model probabilities are estimated to compare the performance of models adopting different contact definitions. Associations between protective behaviours, participant characteristics and setting are explored at the level of individual activities using multiple correspondence analysis and Fisher's exact tests.

Results: Participation in air travel or non-university work activities was associated with a positive asymptomatic SARS-CoV-2 PCR test whereas participation in research or teaching settings was associated with a negative result. Overall, hand washing, social distancing and mask wearing were associated with negative tests but patterns of behaviour varied between settings. Research and teaching settings were associated with mask wearing and hand washing, but not with always being able to socially distance. Interactions during non-university work were not associated with any of the protective behaviours and often occurred in reportedly unventilated spaces. Mask wearing, social distancing and hand washing was less likely to occur during air travel activities.

Conclusions: Research and teaching activities likely presented lower infection risk than other activities undertaken by participating staff and students. University members who travel or work in other settings may pose risk for introducing SARS-CoV-2 infections into a university during term time, particularly during periods where there are restrictions on socialising outside of the household.

Key words: SARS-CoV-2, asymptomatic infection, universities, contact patterns, transmission risk, protective behaviours, mask wearing, social distancing, hand washing

Highlights

- We link asymptomatic PCR SARS-CoV-2 testing to behaviour for members of a UK university.
- Protective behaviours were more often adopted in teaching and research settings.
- During the study non-university work and travel were associated with infection.
- Setting type may be more important than contact numbers when modelling infection risk.
- Infection control in universities is complicated by connections with the community.

1 Introduction

In the 2020/2021 academic year 37% of 18-year-olds in the UK were offered a higher education place [3] and altogether approximately 2.5 million students are registered in higher education in the UK across over 160 providers [14]. Universities provide much of this education; many comprising of order tens of thousands students and staff [14] with highly connected communities through teaching, research, leisure and residential networks. Prior to the emergence of SARS-CoV-2, there was limited data available on the contact networks of university members, nonetheless preliminary modelling studies flagged universities as settings of potential high risk for SARS-CoV-2 transmission [4, 15] Despite mitigations to reduce transmission risk, many universities in the UK [34] experienced outbreaks of SARS-CoV-2 at the beginning of the 2020/2021 academic year [34], some of which may have amplified infection rates in their local community [28]. Students in halls of residence were noted to be at higher risk of experiencing SARS-CoV-2 infection [5], however insight into the risk associated with other activities undertaken by university members is limited.

We interrogate the risk of interactions in settings visited by university members via linking asymptomatic SARS-CoV-2 testing data with quantitative data on social interactions while on and off campus in the week preceding an asymptomatic SARS-CoV-2 PCR test result. For a pathogen with potential for aerosolised and fomite transmission, such as SARS-CoV-2 [20], contacts need not be close or conversational contacts, as typically measured in many contact surveys [e.g. 22]. Furthermore, contact networks may be modified by the adoption of protective behaviours such as mask wearing, social distancing and hand washing. It is thus of interest to examine the role of protective behaviours in concert with a broad definition of contact when surveying the potential transmission risk of a particular activity. We first explore the ability of different individual contact definitions, that variably account for the duration, number of contacts, and presence of extra-household members in each setting to predict asymptomatic SARS-CoV-2 test outcome amongst participating staff and students. We then pool activities across individuals and consider correlations between protective behaviours, setting type and participant characteristics, to examine why interactions in particular settings may present enhanced infection risk.

2 Methods

2.1 Data collection and curation

Participant enrollment was based on a convenience sample of university staff and students returning SARS-CoV-2 positive saliva samples via the Nottingham Asymptomatic Testing (NATS) service. We select a sample of participants testing negative to have a similar distribution of University roles (staff or student), age and gender as the sample of positive participants. Ethical approval was obtained via the University of Nottingham Faculty of Medicine and Health Sciences ethics board (FMHS 96-0920). Participants provided informed consent. A copy of the survey is available upon request.

We divide participants into three groups; those who tested positive (positive), tested negative and had never had a positive test (negative) and those who tested negative but when surveyed had previously tested positive (previous). Unless stated otherwise the analysis is performed on the individuals in the groups positive and negative, minimising any bias due to the impact of SARS-CoV-2 immunity or assumed immunity on susceptibility and behaviour.

Participants were asked to recall information about social interactions and protective behaviours in each activity outside their home undertaken 7 days preceding the receipt of an asymptomatic PCR SARS-CoV-2 test result using a structured interview, offered in person or online. This interview was developed by two psychologists (KV HK) and piloted prior to use. If prompting was needed, participants are encouraged to check their calendars or social media feed. Protective behaviours include whether the participant wore a mask, socially distanced (over 2m away from possible contacts) and cleaned (washed or hand sanitised) their hands before and/or after each activity. Activities are assigned one of twelve settings. These are; abroad/aeroplane, campus other, exercise, hospitality, non-university work, non-private travel, other, research, retail, social, teaching and testing (Supplementary file 1). Survey questions were motivated to capture adherence to pre-July 2021 guidelines for COVID-secure workplaces [7].

Participants are prompted to recall each transition to a new activity and estimate the number of people present (0, 2-5, 5-10, 10-20, 20-50, 50-100, 100+) and duration of the activity. Note that we use the term contact in the broad sense of others present in the same setting, unlike many social contact surveys that assume contacts involve touch or conversation [e.g. 22]. To capture social contact behaviour that could plausibly result in transmission via combinations of droplets, aerosol and/or fomites, we consider seven contact definitions when constructing setting-specific contact measures over the 7 day survey period: participation in the setting, the number of distinct activities, total contacts (the sum of the mid-points of estimated contacts during each activity), total duration of activities, and person-contact-hours (PCH) calculated as the summed product of the midpoint of the estimated contacts and the duration (in hours) of each activity, the total contacts not including the participant's household members and PCH not including the participant's household members. We conservatively set a maximum contact number of 100 for the 100+ option when computing contact measures. We assume that the number of possible household contacts is equivalent to the participant's household size. Therefore, if the number of household contacts is not given for an activity we calculate the non-household contacts and PCH as the difference between the maximum possible household contacts and reported contacts, providing a lower limit on the non-household contacts.

2.2 Setting-specific contact measures predictive of asymptomatic SARS-CoV-2 PCR test result

We use a logistic regression model to regress asymptomatic SARS-CoV-2 test result on contact measures. Due to separation issues in the data (all participants who visited the aeroplane/abroad or non-university work setting tested positive) a Bayesian logistic regres-

sion is performed in *Stan* [30] with a logit link function. As recommended by Gelman et al. [12] priors for the logistic regression coefficients are assumed to follow independent Cauchy distributions centred at 0 and with scale parameter 10 for the constant term and 2.5 for all other coefficients [12]. Prior to fitting the data the binary input (whether a setting was visited) was transformed to have mean 0 and the numeric inputs (all others) are scaled to have mean 0 and standard deviation 0.5 [12].

We assess the significance of regression covariates using Bayes factors. Since the model predictions are sensitive to the prior a support interval (SI) is computed for the coefficient for each setting-specific contact measure [39]. SIs provide information regarding the change in the credibility of values from the prior to the posterior indicating which values of a parameter gain support. Here we present values receiving 'moderate support', with a Bayes factor larger than 3, using the *bayestestR* package [21]. A leave-one-out error analysis is performed on the bounds of the SIs (Supplementary file 3).

We estimate the marginal likelihoods and posterior model probabilities (PMPs) for the logistical model for each set of setting-specific contact measures generated by a contact definition using the *bridgesampling* package [13]. The PMPs are rescaled to sum to 1 across models considered. To examine the support for each contact definition we compute PMPs, averaging over 10 repetitions of the bridge sampling procedure to obtain an empirical estimate of the estimation uncertainty. For the model associated the largest PMP we present posterior predictive checks for the positive and negative groups, and generate an out of sample prediction for test outcome in the previous group.

2.3 Protective behaviours

To examine whether protective behaviours performed during an activity are influenced by the setting or environment we consider the relationship between these behaviours and the university role, gender, age and SARS-CoV-2 test result. As protective behaviours vary between activities even for individuals in the same setting, we pool individual data for this analysis. Each activity is described by nine properties; (i) age of the participant, (ii) gender of the participant, (iii) role (UG, PG, or employed), (iv) test result, (v) setting, (vi) environment (outdoors, ventilated indoors or unventilated indoors), (vii) did the participant wear a mask, (viii) did the participant socially distance at all times and (ix) did the participant use hand sanitiser or wash their hands before and/or after? To assess room ventilation participants were given the example of a room with doors and/or windows open being ventilated.

We perform a multiple correspondence analysis (MCA) in *Rstudio* [27] using the *FactoMineR* package [18] to examine the relationship between these properties. Since responses were not given for all activity properties, the *missMDA* package was used to estimate the number of dimensions for the MCA by leave-one-out cross-validation and impute missing values by cross-validation. Age is denoted a quantitative supplementary variable and gender, role, test result and setting as qualitative supplementary variables. The MCA is performed on the remaining 'active' variables (environment, mask wearing, social distancing and hand washing). Coordinates for the supplementary variables are predicted using the information from this

MCA. The *dimdesc* function is used to determine which categorical variables best describe each dimension and whether age (the continuous variable) is correlated to each dimension. For the quantitative variable, age, correlation coefficients are calculated. For the categorical variables, a univariate anova model is performed for each variable and dimension. An F-test examines whether each variable influences the dimension.

Fisher's exact test is performed on each pair of variables to determine whether they are significantly linked, with p-values corrected for multiple comparisons using the Benjamini–Hochberg method [1]. Results are compared to the Bayesian logistic regression and used to verify consistency of the MCA analysis.

3 Results

Participants are predominantly students between 18 and 30, however staff up to 60 years old are represented. The majority of participants completed the survey online (50, 23 positive and 27 negative), the remaining 12 (9 positive, 3 negative) completed the interviews in person. In all we have data on 447 distinct activities from the 62 participants. There were 20 participants in the positive group, 29 participants in the negative group and 13 in the previous group. The test result dates of the positive group are skewed toward the early period of NATS operation, reflecting the epidemic of self-reported PCR-confirmed SARS-CoV-2 infection within the University [34] (Figure S1a).

The mean and standard deviation of each setting-specific contact measure is provided in Table S1. Retail settings were visited by the largest proportion of participants and had the highest mean non-household contacts, with exercise the most frequently reported activity. Research settings had the highest mean activity duration and mean non-household PCH, with teaching the highest mean PCH. Figure 2 shows the proportion of individuals who participated, mean number of activities, mean total contacts, mean total duration, mean total PCH, mean non-household contacts and mean- non-household PCH for each setting by test result. Of interest, the mean number of contacts across all activities was highest in the negative group, but mean PCH and non-household PCH was higher in the positive group. A Kruskal-Wallis test showed no significant differences between the positive, negative and previous groups for total distinct types of activity ($\chi^2 = 0.17$, df = 2, p = 0.92), number of activities ($\chi^2 = 4.93$, df = 2, p = 0.08), number of contacts ($\chi^2 = 0.76$, df = 2, p = 0.69), duration of activities ($\chi^2 = 1.91$, df = 2, p = 0.38) or PCH ($\chi^2 = 0.44$, df = 2, p = 0.80). There was however a significant difference between the non-household contacts ($\chi^2 = 9.71$, df = 2, p = 0.008) and non-household PCH ($\chi^2 = 6.78$, df = 2, p = 0.03). The mean household size of participants who tested positive was 2.9 (sd = 2.6), whereas the mean household size of participants who tested negative was 3.1 (sd = 2.0).

Protective behaviours reportedly practiced in activities by setting are summarised in Table 1. The percentage of participants who provided answers for whether they wore a mask or socially distanced was the same for all activities except teaching. In teaching settings the participant

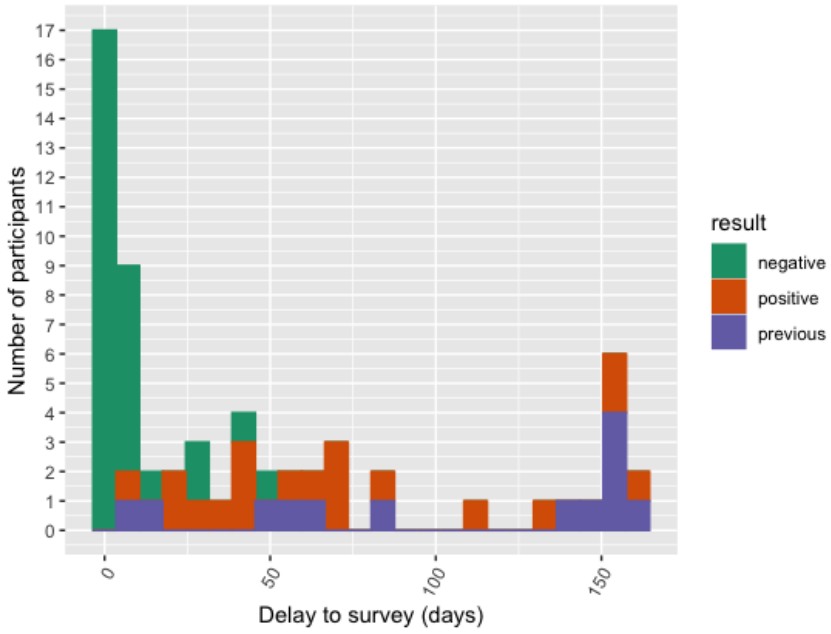

Figure 1: Delays from test result to survey completion by test outcome. Breakdown of test and survey dates by participant age and result is provided in the Supplementary Information .

was only asked about socially distancing. However, for 78% of teaching activities students stated that they wore a mask in the "additional comments" free text field. At the time of this study mask wearing was compulsory (unless exempt) during teaching activities at the university.

3.1 Contact measures predictive of asymptomatic SARS-CoV-2 PCR test result

Table S2 shows the median, minimum and maximum marginal likelihood estimates PMP for each contact definition. The narrow range of the marginal likelihood estimates indicates that the estimation uncertainty is small. The model adopting participation as a contact definition had the most support (PMP = 0.39), followed number of activities (PMP = 0.20) and the non-household PCH (PMP = 0.18). The duration of activities and PCH received less support (0.09 and 0.08, respectively). The models with the least support were the number of contacts and non-household contacts in each setting (PMP = 0.03).

Table 2 gives the coefficient ranges that received moderate support, with a Bayes factor larger than 3, for each co-variate and predictor. We observe that the model with the most support, whether participants entered each setting, has coefficients with moderate evidence which are all positive for the covariates abroad/aeroplane and non-university work and only negative for the covariates research and teaching. Household size, campus other, exercise, hospitality, non-private travel, other, retail, social and testing have support for coefficients with both positive and negative signs.

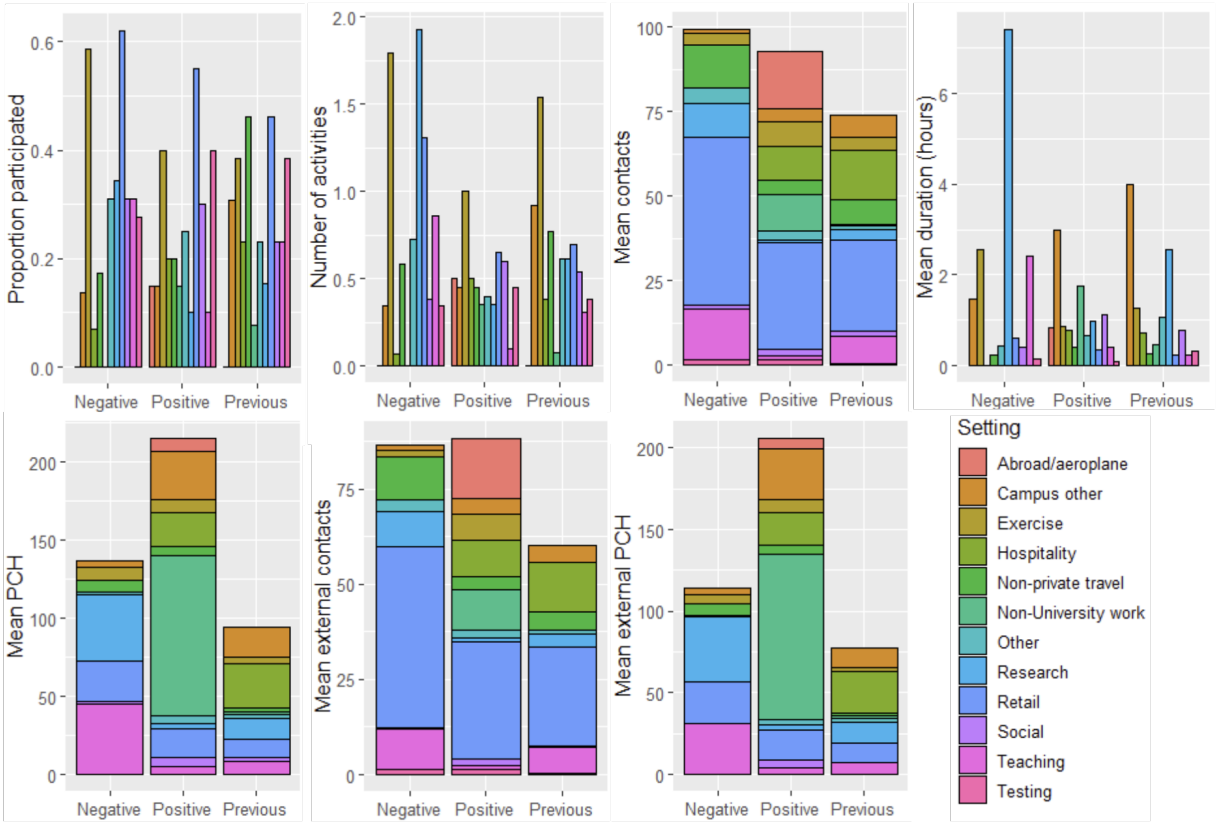

Figure 2: The means of the number of contacts, PCH, duration, non-household contacts and non-household PCH in each setting by test result.

Some coefficients have support intervals (SIs) with the same sign for contact measures other than whether participants entered each setting, however many of these are for models with low support. The model adopting the number of activities as a predictor yields only negative results for exercise, and this also holds when PCH is used as a predictor. When observing coefficients in the support interval for exercise with whether the participants entered each setting we notice this distribution is skewed towards negative values. Other covariates which had SIs with a single sign were non-private travel (negative for PCH and non-household contacts), other (negative for contacts) and social (positive for PCH and negative for non-household contacts), however models with these contact measures as predictors received small PMPs. The model with the largest PMP (based on the participation contact definition) displayed stable SIs across the leave-one-out deletions, but models with low PMPs have large standard deviations in the bounds for the SIs, and signs of the SI bounds are not always consistent with those obtained in the full analysis (Supplementary file 3).

Figure 3 shows the predicted probability of positive asymptomatic PCR SARS-CoV-2 test for the positive, negative and previously positive groups using the logistic regression with predictors indicating participation in an activity in each setting. The means (medians) for the positive, negative and previous groups where 0.60 (1), 0.04 (0) and 0.23 (0) (Table S3).

	Respiratory protection			Hand cleaning				
Setting	Mask	SD	Ans.	No	Before	After	Both	Ans.
Aeroplane/abroad (10)	0	0	70	90	10	10	10	100
Campus other (19)	74	74	100	6	94	94	94	95
Exercise (72)	41	79	99	9	57	91	57	97
Hospitality (12)	17	25	100	17	75	58	50	100
Non-private travel (26)	96	31	100	54	46	46	46	100
Non-university work (7)	100	14	100	100	100	100	100	100
Other (29)	52	72	100	17	62	79	59	100
Research (63)	89	62	97	2	89	98	80	98
Retail (51)	95	39	86	2	63	94	59	100
Social (23)	5	45	87	26	57	74	57	100
Teaching (27)	-	54	96	0	100	96	96	96
Testing (19)	82	82	89	5	89	95	89	100

Table 1: The percentage of activities in each setting where a protective behaviour was performed. Data given are percentage who wore a mask, socially distanced (SD) at all times and when they washed their hands and the percentage of activities which these questions were answered (Ans.). Here individuals who washed their hands both are included in the percentage of people who washed their hands before and after.

Note that a greater proportion of the previous group were students (11/13) compared to the negative group (21/29), however a Fisher's exact test indicated this was not a significant difference at the 5% level.

The Fisher's test analysis is consistent with the results from the logistic regression, with a positive correlation between a positive test result and abroad/aeroplane and non-university work settings and a negative correlation between a positive test result and research and teaching. The Fisher's tests also show a positive correlation between a positive test result and hospitality that was not upheld in the regression analysis, likely due to adjustment for participant interactions in other settings.

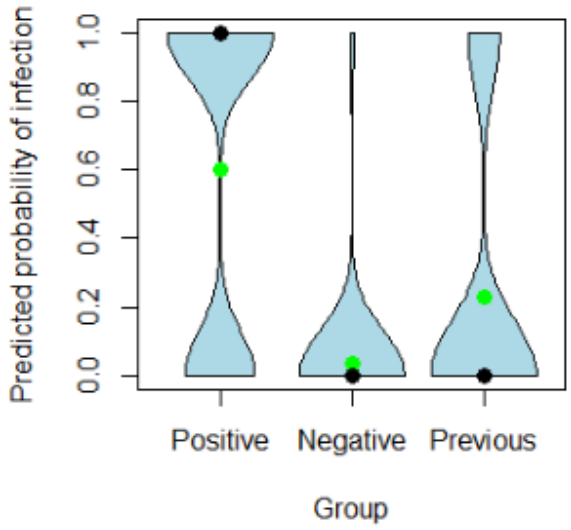

Figure 3: Predicted probabilities of positive asymptomatic PCR SARS-CoV-2 test calculated using the logistic regression with participation in any activity in each setting as the predictive contact measure for each group of participants. Medians are given as a black dot and means are given as a green dot.

Covariate	Participated	Activities	Contacts	Duration	PCH	Contacts*	PCH*
Constant	[-8.12 4.63]	[-8.38 4.51]	[-8.03 4.81]	[-8.54 4.68]	[-8.58 4.68]	[-8.39 4.76]	[-6.87 3.87]
HH size	[-0.85 0.35]	[-0.90 0.30]	[-0.70 0.46]	[-0.92 0.31]	[-0.65 0.51]	[-0.71 0.45]	[-0.65 0.52]
Abroad/aeroplane	[1.91 3.19]†	-	-	-	-	-	-
Campus other	[-0.68 0.57]	[-0.35 0.83]	[-0.27 0.92]	[-0.27 0.91]	[-0.11 1.09]	[-0.21 0.98]	[-0.12 1.08]
Exercise	[-1.22 0.04]	[-2.00 -0.51]†	[-0.29 1.00]	[-1.61 -0.25]†	[-0.95 0.31]	[-0.23 1.05]	[-0.74 0.42]
Hospitality	[-0.92 0.26]	[-0.28 0.81]	-	[-0.03 1.25]	-	-	-
Non-university work	[2.14 5.38]†	[0.77 2.45]†	[-0.02 1.18]	[0.21 1.50]†	[0.20 1.04]†	[-0.03 1.18]	[0.25 1.13]†
Non-private travel	[-1.26 0.01]	[-1.25 0.10]	[-1.20 0.06]	[-1.06 0.25]	[-1.10 -0.24]†	[-1.13 0.04]	[-1.09 -0.23]†
Other	[-0.54 0.74]	[-0.89 0.32]	[-1.29 -0.05]†	[-1.15 0.16]	-	[-1.10 0.08]	-
Research	[-1.80 -0.32]†	[-1.37 0.03]	[-1.28 -0.05]†	[-1.31 -0.05]†	[-1.32 -0.06]†	[-1.18 0.03]	[-1.25 -0.00]†
Retail	[-0.37 0.94]	[-1.18 0.04]	[-0.89 0.25]	[-0.89 0.30]	[-0.74 0.46]	[-1.07 0.19]	[-0.66 0.48]
Social	[-0.35 0.91]	[-0.04 1.29]	[-0.91 0.25]	[-0.05 1.31]	[0.24 1.17]†	[-1.05 -0.01]†	-
Teaching	[-2.45 -0.78]†	[-1.70 -0.33]†	[-1.35 0.02]	[-1.38 -0.04]†	[-1.34 0.06]	[-1.22 0.03]	[-1.41 -0.41]†
Testing	[-0.96 0.36]	[-0.64 0.67]	[-0.53 0.71]	[-0.82 0.43]	[-0.72 0.51]	[-0.33 0.87]	[-0.61 0.61]

Table 2: Support intervals (SIs) of the constant and co-variate coefficients for each predictor. SIs give ranges of parameters with a Bayes factor larger than or equal to 3, interpreted as moderate support. * = predictors include only external non-household contacts. † = SIs where all supported values have the same sign. Results are given to 2 decimal places.

3.2 Protective behaviours

Altogether, four dimensions are required to explain 70.9% of the variance in the MCA analysis (Figure S2). Figure 4 shows how the active and qualitative variables relate to the first two dimensions of the MCA, and provides evidence that patterns of protective behaviour differ between settings. Significant components for the first four dimensions are provided and results from the Fisher's exact tests are provided in Supplementary file 4.

Figure 4: Graph visualising the coordinates of each variable categories in dimensions 1 and 2. The distance between any points gives a measure of their similarity. Supplementary variables are shown in green and active variables are shown in red.

Fisher's tests showed positive test results were positively correlated with no hand washing and females, and were negatively correlated with mask wearing, washing hands after and maintaining social distancing at all times.

Of the activities on campus, teaching and retail are associated with wearing a mask and washing hands both before and after. PGs activities are positively correlated with indoor unventilated environments. They are also more likely to socially distance and wash their hands. Being classified as staff is negatively correlated with hand washing.

Outside of campus-based activities we find that retail was associated with mask wearing and

washing hands before and after an activity. Masks are unlikely to be worn in social settings and outdoors. The MCA analysis shows an association between the abroad/aeroplane setting and not cleaning hands before or after travel. The Fisher's analysis indicates a negative correlation of air travel with wearing a mask and social distancing. Participants in the study that were staff were more likely to participate in travel activities.

In our sample only the abroad/aeroplane setting was solely visited by females, which likely contributed to the significance of females receiving a positive result. Females (who make up 61% of our total sample) were also more likely to be in an UG role (75% of UGs female) than PG (50% of PGs) or employed (53% of staff)). This may explain the negative correlation between research setting and females (which participating in, according to the logistic regression, reduced the probability of positive asymptomatic SARS-CoV-2 PCR test). It is therefore plausible that female and positive test result are positively correlated due to their role at the university and the relatively small sample size.

4 Discussion

We have presented linked asymptomatic SARS-CoV-2 testing and social contact data for a UK university collected from October 2020 to March 2021, after the initial surge in infections at the beginning of the 2020/2021 academic year [34]. Restrictions on mixing outside of your household/bubble and educational activities were in place throughout the study period, with particularly stringent rules about social contact with those outside of your household in place during tier 3 restrictions (30 October 2020–5 November 2020), and the second (5 November 2020–2 December 2020) and third (6 January 2021–8 March 2021) national lockdowns [35, 36]. Strict social distancing measures were in place on campus throughout the study period. Teaching was undertaken in a hybrid (online/in-person) manner to accommodate social distancing. During the third national lockdown only students on a limited selection of courses (Medicine Dentistry, Health Social Care including Nursing Midwifery, Physiotherapy and Veterinary Science, Education, and Social Work) were permitted to travel to campus without exemption. Research staff were asked to work from home whenever possible [37].

Within the cohort studied, participating in air travel (and holiday activities) and non-university work increase the probability of an asymptomatic SARS-CoV-2 PCR test result. This is consistent with evidence that workers with public facing roles are at higher risk of infection (see [25] and references therein, [32]). Students or staff in part-time work in other settings could mediate spillover between the University and community. Encouraging vaccination of these groups could be particularly important for mitigating the risk of university outbreaks in periods of restrictions.
Participating in research and teaching activities at the university was associated with a lower risk of a positive asymptomatic SARS-CoV-2 PCR test result, however no association was found with participation in other activities on campus. Although our data does not permit estimate of the risk attributable to any activity or setting, our results are consistent with teaching and research activities having lower infection risk than other activities noted by

participants. Participation in a setting was the best-performing contact measure, followed by counts of the number of activities in each setting and the non-household PCH (each for 7 days preceding test result). Out of sample model predictions for the group previously testing positive yielded expected test positivity between that for the negative and positive groups across all contact definitions.

Indications that non-household PCH and duration of activities provide a better predictor of SARS-CoV-2 PCR test status than contacts or PCH provides tentative evidence for the role of contact duration in infection risk, also reported by Thompson et al. [33]. There are several potential reasons for the difference in model performance between contact definitions, and in particular the superior performance of participation in a setting rather than the definitions capturing contact numbers. Infected people can remain PCR positive for up to 3 weeks following exposure [29] depending on the sensitivity of the saliva assay [31] and therefore we can not guarantee we have surveyed participant behaviour during the period of exposure. Additionally, contacts in each activity were assumed to extend for the duration of an activity. We have not collected data on repeated contacts, and it is plausible that the participation contact definition is preferred because of this. Predictors that capture the nature and duration of every contact in each activity may perform better. While collecting more detailed contact data was considered prohibitive from a recall perspective in our retrospective survey, similar contact diary studies have been piloted [2], and it would be of interest to link these to repeated asymptomatic test outcome data for SARS-CoV-2 or other respiratory viruses in larger studies.

We adjusted for participant household size in our regression, but did not find a significant effect. Other studies examining the risk associated with household size have been mixed, with an analysis of setting-dependent transmission risk not identifying household size as significant [33], but recent SARS-CoV-2 prevalence higher in larger households [24]. Analysis of contact patterns in another UK university suggested that extra-household contacts were higher amongst those living in smaller households [23], and it is plausible that such an effect could offset the enhanced risk of importation into a larger student households. Outbreaks in halls of residence may have been more strongly influenced by hall than household size [28] and it is possible that there are other risks associated with residential contacts in this setting that we are not capturing.

Positive asymptomatic SARS-CoV-2 PCR test results were negatively associated with mask wearing, social distancing and hand washing, as reported in another case-control study of asymptomatically infected contacts of SARS-CoV-2 cases [8]. Our MCA highlighted differential adoption of protective measures between settings and suggests that protective behaviours can be different in university and non-university activities. Teaching and research setting may be lower risk (despite similar or larger mean contact measures across contact definitions) as they were associated with mask wearing and hand washing. Although mask wearing and hand washing was practiced uniformly in non-University work settings, this was less likely to be in a ventilated space with complete adherence to social distancing than other settings. Other studies suggest that adoption of protective behaviours is also determined by psychological factors [10, 40, 41], which could explain some of the variability between

participant behaviour. Measuring the prevalence of micro-distancing behaviour as well as macro-distancing behaviour now widely captured by mobility patterns has shown utility in estimating the effective reproduction number in low-prevalence settings (see, e.g., [9]) and may also aid in parameterising agent-based simulation of transmission [19].

Our results come with a number of caveats. Our sample was chosen based on consenting positive cases and may not be representative of all users of the NATS or the wider University population. A greater proportion of positive cases in our study were from periods with lower levels of restriction which could generate time-varying confounding in our analysis. A larger sample would likely allow for adjustment for this and other potential confounders, and potentially provide statistical power to include setting-specific protective behaviours in the regression model. As discussed elsewhere [26], the opportunities for contact and transmission depend on community prevalence and the social restrictions in place. Data for this study was collected over a period during which there were strong (albeit changing) restrictions on permitted social and travel activities, which may partly explain the absence of a significant effect of social interactions on risk of obtaining a positive test result as reported in other studies [17]. Similarly, occupancy on campus was low during the study with much teaching online, and we expect the relative risk of activities in different settings will change depending on how university and national policies, and the behavioural response to these, evolve. Furthermore, participants were surveyed at a time when the circulating SARS-CoV-2 strain was either phenotypically akin to the original Wuhan strain, or the alpha variant of concern, and it is plausible that different patterns of risk would be observed for delta or other variants with different infectiousness profiles.

Although the structured interview adopted aims to optimise recall of social contact behaviour, the limitations of recalling such details accurately are well documented [e.g. 11]. Participants who received positive test results could be experiencing stress/anxiety that may influence their ability to recall events [11]. Others have suggested that recall bias could act in the opposite direction, with SARS-CoV-2 positive participants more likely to recall possible contact events [6]. The significant delays between test result and survey (Figure S1a) may also influence recall ability [16]. Our findings relate to a highly educated population, a characteristic that has been associated with adopting protective behaviours [38]. Despite likely ready access to PPE and other resources enabling protective behaviour, protective behaviours were not uniformly reported amongst participants. In future work we plan to analyse free text responses noting the challenges in adhering to social distancing guidelines in university and other community settings, and explore patterns of contact and protective behaviour leading to decisions to undertake voluntary asymptomatic testing.

5 Acknowledgements

The authors thank the Nottingham Asymptomatic Testing Service.

6 Author contributions

EF: data curation, formal analysis, investigation, methodology, software, validation, visualisation, writing - original draft, writing - reviewing and editing. KB: conceptualisation, funding acquisition, investigation, methodology, software, supervision, writing - original draft, writing - reviewing and editing. RJ: conceptualisation, investigation, methodology, project administration, writing - reviewing and editing. GF: conceptualisation, funding acquisition, data analysis, writing - reviewing and editing. HK: conceptualisation, methodology, project administration, writing - reviewing and editing. KV: conceptualisation, funding acquisition, investigation, methodology, resources, software, supervision, writing - reviewing and editing.

7 Financial support

This work was funded by the University of Nottingham. EF. acknowledges support via the Nottingham BBSRC Doctoral Training Partnership. KB acknowledges support of a University of Nottingham Anne McLaren Fellowship.

References

[1] Benjamini, Y. and Hochberg, Y. [1995], 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *J R Stat Soc Series B Stat Methodol* **57**(1), 289–300.

[2] Bolton, K. J., McCaw, J. M., Forbes, K., Nathan, P., Robins, G., Pattison, P., Nolan, T. and McVernon, J. [2012], 'Influence of contact definitions in assessment of the relative importance of social settings in disease transmission risk', *PloS one* **7**(2), e30893.

[3] Bolton, P [2021], *Higher education student numbers.* Accessed: 2021-07-29.
URL: *https://researchbriefings.files.parliament.uk/documents/CBP-7857/CBP-7857.pdf*

[4] Brooks-Pollock, E., Christensen, H., Trickey, A., Hemani, G., Nixon, E., Thomas, A., Turner, K., Finn, A., Hickman, M., Relton, C. and Danon, L. [2021], 'High COVID-19 transmission potential associated with re-opening universities can be mitigated with layered interventions', *Nat Commun* **12**, 5017.

[5] Children's Task and Finish Group [2021], *Paper on Higher Education Settings.* Accessed: 2021-07-29.
URL: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/963387/S1103_Children_s_Task_and_finish_Group_Paper_on_Higher_Education_Settings_.pdf*

[6] Delgado-Rodríguez, M. and Llorca, J. [2004], 'Bias', *J Epidemiol Community Health* **58**(8), 635–641.

[7] Department for Business, Energy & Industrial Strategy and Department for Digital, Culture, Media & Sport [2020], *Working safely during coronavirus (COVID-19)*. Accessed: 2020-11-30.
URL: *https://www.gov.uk/guidance/working-safely-during-coronavirus-covid-19*

[8] Doung-Ngern, P., Suphanchaimat, R., Panjangampatthana, A., Janekrongtham, C., Ruampoom, D., Daochaeng, N., Eungkanit, N., Pisitpayat, N., Srisong, N., Yasopa, O., Plernprom, P., Promduangsi, P., Kumphon, P., Suangtho, P., Watakulsin, P., Chaiya, S., Kripattanapong, S., Chantian, T., Bloss, E., Namwat, C. and Limmathurotsakul, D. [2020], 'Case-control study of use of personal protective measures and risk for SARS-CoV 2 infection, Thailand', *Emerg Inf Dis* **26**(11), 2607.

[9] *Estimating temporal variation in transmission of SARS-CoV-2 and physical distancing behaviour in Australia, Technical Report 17 July 2020* [2020].
URL: *https://www.doherty.edu.au/uploads/content_doc/Technical_report_4_update_29July2020.pdf*

[10] Faasse, K. and Newby, J. [2020], 'Public perceptions of COVID-19 in Australia: perceived risk, knowledge, health-protective behaviors, and vaccine intentions', *Front Psychol* **11**, 551004.

[11] Garry, M., Hope, L., Zajac, R., Verrall, A. J. and Robertson, J. M. [2021], 'Contact Tracing: A Memory Task With Consequences for Public Health', *Perspect Psychol Sci* **16**(1), 175–187.

[12] Gelman, A., Jakulin, A., Pittau, M. G. and Su, Y.-S. [2008], 'A weakly informative default prior distribution for logistic and other regression models', *Ann Appl Stat* **2**(4), 1360–1383.

[13] Gronau, Q. F., Singmann, H., Forster, J. J., Wagenmakers, E.-J., The JASP Team, Guo, J., Gabry, J., Goodrich, B., Mulder, K. and de Valpine, P. [2020], *bridgesampling: Bridge Sampling for Marginal Likelihoods and Bayes Factors*.
URL: *https://github.com/quentingronau/bridgesampling*

[14] Higher Education Statistics Agency [2021], *Open Data and Official Statistics*. Accessed: 2021-07-29.
URL: *https://www.hesa.ac.uk/data-and-analysis*

[15] Hill, E. M., Atkins, B. D., Keeling, M. J., Tildesley, M. J. and Dyson, L. [2021], 'Modelling SARS-CoV-2 transmission in a UK university setting', *Epidemics* **36**, 100476.

[16] Hipp, L., Bünning, M., Munnes, S. and Sauermann, A. [2020], 'Problems and pitfalls of retrospective survey questions in COVID-19 studies', *Surv Res Methods* **14**(2), 109–1145.

[17] Hobbs, C. V., Martin, L. M., Kim, S. S., Kirmse, B. M., Haynie, L., McGraw, S., Byers, P., Taylor, K. G., Patel, M. M., Flannery, B. and CDC COVID-19 Response Team [2020], 'Factors Associated with Positive SARS-CoV-2 Test Results in Outpatient

Health Facilities and Emergency Departments Among Children and Adolescents Aged< 18 Years—Mississippi, September–November 2020', *MMWR Morb Mortal Wkly Rep* **69**(50), 1925–1929.

[18] Husson, F., Josse, J., Le, S. and Mazet, J. [2008], *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining.*
URL: *http://http://factominer.free.fr/*

[19] Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., Núñez, R. C., Cohen, J. A., Selvaraj, P., Hagedorn, B., George, L., Jastrzebski, M., S, I. A., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S. L., Bennette, C. S., Wagner, B. G., Chang, S. T., Oron, A. P., Wenger, E. A., Panovska-Griffiths, J., Famulare, M. and Klein, D. J. [2021], 'Covasim: an agent-based model of COVID-19 dynamics and interventions', *PLOS Comput Biol* **17**(7), e1009149.

[20] Klompas, M., Baker, M. A. and Rhee, C. [2020], 'Airborne Transmission of SARS-CoV-2: Theoretical Considerations and Available Evidence', *JAMA* **324**, 441–442.

[21] Makowski, D., Lüdecke, D., Ben-Shachar, M. S., Patil, I., Wilson, M. D., Bürkner, P.-C., Mahr, T., Singmann, H., Gronau, Q. F. and Crawley, S. [2019], *bayestestR: Understand and Describe Bayesian Models and Posterior Distributions.*
URL: *https://easystats.github.io/bayestestR/*

[22] Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M. and Edmunds, W. J. [2008], 'Social contacts and mixing patterns relevant to the spread of infectious diseases', *PLoS Med* **5**(3), e74.

[23] Nixon, E., Trickey, A., Christensen, H., Finn, A., Thomas, A., Relton, C., Montgomery, C., Hemani, G., Metz, J., Walker, J. G., Turner, K., Kwiatkowska, R., Sauchelli, S., Danon, L. and Brooks-Pollock, E. [2021], 'Contacts and behaviours of university students during the COVID-19 pandemic at the start of the 2020/2021 academic year', *Sci Rep* **11**(1), 11728.

[24] ONS [2021], *Coronavirus (COVID-19) Infection Survey technical article: analysis of populations in the UK by risk of testing positive for COVID-19, September 2021.*
URL: *https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/analysisofpopulationsintheukbyriskoftestingpositiveforcovid19september2021*

[25] PHE Transmission Group [2020], *PHE: Factors contributing to risk of SARS-CoV-2 transmission in various settings, 26 November 2020.*
URL: *https://www.gov.uk/government/publications/phe-factors-contributing-to-risk-of-sars-cov2-transmission-in-various-settings-26-november-2020*

[26] Royal Society SET-C [2020], 'SARS-CoV-2: Where do people acquire infection and 'who infects whom'?'.

URL: *https://royalsociety.org/-/media/policy/projects/set-c/set-c-transmission-paper.pdf*

[27] RStudio Team [2020], *Rstudio: Integrated Development Environment for R*, RStudio, PBC., Boston, MA.
URL: *http://www.rstudio.com/*

[28] *SARS-CoV-2 infection in UK university students: lessons from September-December 2020 and modelling insights for future student return* [n.d.].

[29] Sethuraman, N., Jeremiah, S. S. and Ryo, A. [2020], 'Interpreting diagnostic tests for SARS-CoV-2', *JAMA* **323**(22), 2249–2251.

[30] Stan [2019], *Stan user's guide version 2.27.*
URL: *https://mc-stan.org/docs/2_27/stan-users-guide/index.html*

[31] Teo, A. K. J., Choudhury, Y., Tan, I. B., Cher, C. Y., Chew, S. H., Wan, Z. Y., Cheng, L. T. E., Oon, L. L. E., Tan, M. H., Chan, K. S. and Hsu, L. Y. [2021], 'Saliva is more sensitive than nasopharyngeal or nasal swabs for diagnosis of asymptomatic and mild COVID-19 infection', *Sci Rep* **11**(1), 3134.

[32] Thomas, D. R., Fina, L. H., Adamson, J. P., Sawyer, C., Jones, A., Nnoaham, K., Barrasa, A., Shankar, A. G. and Williams, C. J. [2021], 'Social, demographic and behavioural determinants of SARS-CoV-2 infection: A case-control study carried out during mass community testing of asymptomatic individuals in South Wales, December 2020', *medRxiv* .
URL: *https://www.medrxiv.org/content/early/2021/04/09/2021.04.06.21253465*

[33] Thompson, H. A., Mousa, A., Dighe, A., Fu, H., Arnedo-Pena, A., Barrett, P., Bellido-Blasco, J., Bi, Q., Caputi, A., Chaw, L., Maria, L. D., Hoffmann, M., Mahapure, K., Ng, K., Raghuram, J., Singh, G., Soman, B., Soriano, V., Valent, F., Vimercati, L., Wee, L. E., Wong, J., Ghani, A. C. and Ferguson, N. M. [2021], 'Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) setting-specific transmission rates: a systematic review and meta-analysis', *Clin Infect Dis* **73**, e754–e764.

[34] UCU [2020], *Monitoring cases of Covid-19 in UK higher and further education.* Accessed: 2020-11-15.
URL: *https://www.ucu.org.uk/covid-dashboard*

[35] UK Government [2020], *The Health Protection (Coronavirus, Restrictions) (England) Regulations 2020.* Accessed: 2021-05-22.
URL: *https://www.legislation.gov.uk/uksi/2020/350/regulation/6/2020-03-26*

[36] UK Government [2021], *The Health Protection (Coronavirus, Restrictions) (England) Regulations 2021.* Accessed: 2021-05-22.
URL: *https://www.legislation.gov.uk/uksi/2021/364/contents/made*

[37] University of Nottingham [2021], 'Latest updates and news'.
URL: *https://www.nottingham.ac.uk/coronavirus/communications/archive-of-communications.aspx*

[38] Vally, Z. [2020], 'Public perceptions, anxiety and the perceived efficacy of health-protective behaviours to mitigate the spread of the SARS-Cov-2/COVID-19 pandemic', *Public Health* **187**, 67–73.

[39] Wagenmakers, E.-J., Gronau, Q. F., Dablander, F. and Etz, A. [2020], 'The support interval', *Erkenntnis* .

[40] Wise, T., Zbozinek, T. D., Michelini, G., Hagan, C. C. and Mobbs, D. [2020], 'Changes in risk perception and self-reported protective behaviour during the first week of the covid-19 pandemic in the united states', *R Soc Open Sci* **7**(9), 200742.

[41] Zickfeld, J. H., Schubert, T. W., Herting, A. K., Grahe, J. and Faasse, K. [2020], 'Correlates of health-protective behavior during the initial days of the COVID-19 outbreak in Norway', *Front Psychol* **11**, 564083.

Bibliography

BN Fields, DM Knipe, and PM Howley. *Fields virology*. Vol. 2 Wolters Kluwer Health. Lippincott Williams & Wilkins Philadelphia:, 2013.

M Hemida, M Alhammadi, A Daleb, and A Alnaeem. Molecular and serological surveillance of African horse sickness virus in Eastern and Central Saudi Arabia. *Rev Sci Tech*, 36(3):889–898, 2017.

PS Mellor and C Hamblin. African horse sickness. *Vet Res*, 35(4):445–466, 2004.

S Carpenter, PS Mellor, AG Fall, C Garros, and GJ Venter. African horse sickness virus: history, transmission, and current status. *Annu Rev Entomol*, 62:343–358, 2017.

J Castillo-Olivares. African horse sickness in Thailand: Challenges of controlling an outbreak by vaccination. *Equine Vet J*, 51(1):9–14, 2021. doi: 10.1111/evj.13353.

G Lu, J Pan, J Ou, R Shao, X Hu, C Wang, and S Li. African horse sickness: Its emergence in Thailand and potential threat to other Asian countries. *Transbound Emerg Dis*, 67 (5):1751–1753, 2020. doi: 10.1111/tbed.13625.

MW Henning. *Animal diseases in South Africa*. South Africa: Central News Agency, Ltd., 1949.

TB Bayley. *Notes on the Horse-sickness at the Cape of Good Hope, in 1854-'55: Compiled, by Permission of His Excellency the Governor, from Official Documents*. S. Solomon & Company, Steam Print. Office, 1856.

M Anwar and MAA Qureshi. Control and eradication of African horse sickness in Pakistan. *Dev Biol*, 199:225–228, 1973.

PG Howell. The 1960 epizootic of African horse sickness in the Middle East and SW Asia. *J S Afr Vet Assoc*, 31(3):329–334, 1960. doi: 10.10520/AJA00382809_817.

J Lubroth. African Horse sickness and the epizootic in Spain, 1987. In *P Annu Conv Am Equin*, 1988.

GM Thompson, S Jess, and AK Murchie. A review of African horse sickness and its implications for Ireland. *Ir Vet J*, 65(1):9, 2012.

M Portas, FS Boinas, J Oliveira, E Sousa, and P Rawlings. African horse sickness in Portugal: a successful eradication programme. *Epidemiol Infect*, 123(2):337–346, 1999.

M Rodriguez, H Hooghuis, and MA Castano. African horse sickness in Spain. *Vet Microbiol*, 33(1-4):129–142, 1992.

PS Mellor. African horse sickness: transmission and epidemiology. *Vet Res*, 24(2):199–212, 1993.

M Baylis, HEL Hasnaoui, H Bouayoune, J Touti, and PS Mellor. The spatial and seasonal distribution of African horse sickness and its potential Culicoides vectors in Morocco. *Med Vet Entomol*, 11(3):203–212, 1997. doi: 10.1111/j.1365-2915.1997.tb00397.x.

ND Diouf, E Etter, MM Lo, M Lo, and AJ Akakpo. Outbreaks of African horse sickness in Senegal, and methods of control of the 2007 epidemic. *Vet Rec*, 172(6):152–152, 2013. doi: 10.1136/vr.101083.

N Aklilu, C Batten, E Gelaye, S Jenberie, G Ayelet, A Wilson, A Belay, Y Asfaw, C Oura, S Maan, K Bachanek-Bankowska, and PPC Mertens. African Horse Sickness Outbreaks Caused by Multiple Virus Types in Ethiopia. *Transbound Emerg Dis*, 61(2):185–192, 2014. doi: 10.1111/tbed.12024.

SL Li, JP Messina, OG Pybus, MUG Kraemer, and L Gardner. A review of models applied to the geographic spread of Zika virus. *Trans R Soc Trop Med Hyg*, 115(9):956–964, 2021.

ARW Elbers, CJ Koenraadt, and R Meiswinkel. Mosquitoes and *Culicoides* biting midges: vector range and the influence of climate change. *Rev Sci Tech*, 34(1):123–137, 2015.

EJ Wittmann and M Baylis. Climate change: effects on *Culicoides*-transmitted viruses and implications for the UK. *Vet J*, 160(2):107–117, 2000.

H Yang, W Gu, Z Li, L Zhang, D Liao, J Song, B Shi, J Hasimu, Z Li, Z Yang, Q Zhong, and H Li. Novel putative Bluetongue virus serotype 29 isolated from inapparently infected goat in Xinjiang of China. *Transbound Emerg Dis*, 68(4):2543–2555, 2021.

The Pirbright Institute. Bluetongue virus, 2019. URL `https://www.pirbright.ac.uk/bluetongue`. Accessed: 2019-09-09.

K Polydorou. The 1977 outbreak of bluetongue in Cyprus. *Trop Anim Health Prod*, 10(1):229–232, 1978.

P Roy. Bluetongue viruses. *Desk Encyclopedia of Animal and Bacterial Virology*, pages 43–50, 2008.

J-F Toussaint, F Vandenbussche, J Mast, L De Meester, N Goris, W Van Dessel, E Vanopdenbosche, Kerkhofs, K De Clercq, S Zientara, C Saileau, G Czaplicki, G Depoorter, and J-M Dochy. Bluetongue in northern Europe. *Vet Rec*, 159(10):327, 2006.

F Landeg. Bluetongue outbreak in the UK. *Vet Rec*, 161(15):534–535, 2007.

J Rushton and N Lyons. Economic impact of Bluetongue: a review of the effects on production. *Vet Ital*, 51(4):401–406, 2015.

J Gethmann, C Probst, and FJ Conraths. Economic Impact of a Bluetongue Serotype 8 Epidemic in Germany. *Front Vet Sci*, 7:65, 2020.

K Wernike, F Conraths, G Zanella, H Granzow, K Gache, H Schirrmeier, S Valas, C Staubach, P Marianneau, F Kraatz, D Höreth-Böntgen, I Reimann, S Zientara, and M Beer. Schmallenberg virus–two years of experiences. *Prev Vet Med*, 116(4):423–434, 2014.

K Wernike and M Beer. Schmallenberg virus: To vaccinate, or not to vaccinate? *Vaccines*, 8(2):287, 2020a.

ÁB Collins, ML Doherty, DJ Barrett, and JF Mee. Schmallenberg virus: a systematic international literature review (2011-2019) from an Irish perspective. *Ir Vet J*, 72(1): 1–22, 2019.

K Wernike and M Beer. Re-circulation of Schmallenberg virus, Germany, 2019. *Transbound Emerg Dis*, 67(6):2290–2295, 2020b.

M Barba, EL Fairbanks, and JM Daly. Equine viral encephalitis: prevalence, impact, and management strategies. *Vet Med (Auckl)*, 10:99–110, 2019.

MT Long. West Nile virus and equine encephalitis viruses: new perspectives. *Vet Clin North Am Equine Pract*, 30(3):523–542, 2014.

B Kumar, A Manuja, BR Gulati, N Virmani, and BN Tripathi. Zoonotic viral diseases of equines and their impact on human and animal health. *Open Virol J*, 12:80–98, 2018.

N Zhu, D Zhang, W Wang, X Li, B Yang, J Song, X Zhao, B Huang, W Shi, R Lu, Peihua Niu, F Zhan, X Ma, D Wand, W Xu, G Wu, W Gao, GF Tan, and China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*, 382:727–733, 2020.

G Dropkin. COVID-19 UK lockdown forecasts and R0. *Front Public Health*, 8:256, 2020.

EM Hill, BD Atkins, MJ Keeling, M Tildesley, and L Dyson. Modelling SARS-CoV-2 transmission in a UK university setting. *Epidemics*, 36:100476, 2021.

E Brooks-Pollock, H Christensen, A Trickey, G Hemani, E Nixon, AC Thomas, K Turner, A Finn, M Hickman, C Relton, and L Danon. High COVID-19 transmission potential associated with re-opening universities can be mitigated with layered interventions. *Nat Commun*, 12(1):5017, 2021.

Task and Finish Group on Higher Education/Further Education. *Principles for managing SARS-CoV-2 transmission associated with higher education, 3 September 2020*, 2020. URL `https://www.gov.uk/government/publications/principles-for-managing-sars-cov-2-transmission-associated-with-higher-education-3-september-2020`. Accessed: 2021-02-05.

UCU. *Monitoring cases of Covid-19 in UK higher and further education*, 2020. URL `https://www.ucu.org.uk/covid-dashboard`. Accessed: 2020-11-15.

MC Chubb and KH Jacobsen. Mathematical modeling and the epidemiological research process. *Eur J Epidemiol*, 25(1):13–19, 2010.

HW Hethcote. The mathematics of infectious diseases. *SIAM Rev Soc Ind Appl Math*, 42 (4):599–653, 2000.

GD Pullinger, M Guimerà Busquets, K Nomikou, M Boyce, H Attoui, and PPC Mertens. Identification of the genome segments of bluetongue virus serotype 26 (Isolate KUW2010/02) that restrict replication in a culicoides sonorensis cell line (KC Cells). *PLoS One*, 11(2):e0149709, 2016.

K Dietz. The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res*, 2(1):23–41, 1993.

CM Guldberg and P Waage. Studies concerning affinity, 1986.

G Macdonald. *The epidemiology and control of malaria*. Oxford University Press, 1957.

JA González, FJ Rodríguez-Cortés, O Cronie, and J Mateu. Spatio-temporal point process statistics: a review. *Spat Stat*, 18:505–544, 2016a.

R Owers and K Meldrum. The demise of the UK's National Equine Database. *Equine Vet J*, 45(3):269–269, 2013.

JD Grewar, JL Kotze, BJ Parker, LS van Helden, and CT Weyer. An entry risk assessment of African horse sickness virus into the controlled area of South Africa through the legal movement of equids. *PLoS One*, 16(5):e0252117, 2021. doi: 10.1371/journal.pone.0252117.

Food Department for Environment and Rural Affairs. *African Horse Sickness in horses in South Africa*, 2016. URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/517121/african-horse-sickness-south-africa-april2016.pdf`. Accessed: 2021-09-22.

LA Boden, TDH Parkin, J Yates, D Mellor, and RR Kao. Summary of current knowledge of the size and spatial distribution of the horse population within Great Britain. *BMC Vet Res*, 8(1):43, 2012.

ME Ensminger. *Horses and Horsemanship (Animal Agriculture Series)*. Interstate Publishers, Danville, Illinois, USA, 1990.

CA Robin, G Lo Iacono, S Gubbins, JLN Wood, and JR Newton. The accuracy of the National Equine Database in relation to vector-borne disease risk modelling of horses in Geat Britain. *Equine Vet J*, 45(3):302–308, 2013.

CA Robin, CE Wylie, JLN Wood, and JR Newton. Making use of equine population demography for disease control purposes: preliminary observations on the difficulties of counting and locating horses in Great Britain. *Equine Vet J*, 43(3):372–375, 2011.

LA Boden, TDH Parkin, J Yates, D Mellor, and RR Kao. An online survey of horse-owners in Great Britain. *BMC Vet Res*, 9(1):1–11, 2013.

G Lo Iacono, CA Robin, JR Newton, S Gubbins, and JLN Wood. Where are the horses? With the sheep or cows? Uncertain host location, vector-feeding preferences and the risk of African horse sickness transmission in Great Britain. *J R Soc Interface*, 10(83):20130194, 2013.

RM Fuller, GM Smith, JM Sanderson, RA Hill, and AG Thomson. The UK Land Cover Map 2000: construction of a parcel-based vector map from satellite images. *Cartogr J*, 39(1):15–25, 2002.

Estimation of indicators on the number of equines and the jobs they generate in France, author=Dornier, X, journal=Institut français du cheval et de l'équitation, pages=251–254, year=2010.

H Farchati, A Merlin, M Saussac, X Dornier, M Dhollande, D Garon, J Tapprest, and C Sala. Is the French SIRE equine information system a good basis for surveillance and epidemiological research? Quality assessment using two surveys. *Res Vet Sci*, 134:96–101, 2021.

KS Vanderman, NA Dreschel, AM Swinker, DM Kniffen, RB Radhakrishna, JR Werner, and EA Jedrzejewski. Equine veterinarians' and health care professionals' concerns related to the implementation of the National Equine Identification System. *J Equine Vet Sci*, 29(12):823–827, 2009.

AM Swinker, KS Vanderman, DM Kniffen, BE Gill, AL Graeff, MI Gerber, and KA Vines. A Survey Overview of United States Equine Show and Event Managers on the National Equine Identification System. *Prof Anim Sci*, 25(6):774–778, 2009.

W Hartig, H Houe, and PH Andersen. Monitoring of equine health in Denmark: the importance, purpose, research areas and content of a future database. *Prev Vet Med*, 109(1-2):92–105, 2013a.

W Hartig, H Houe, and PH Andersen. Monitoring of equine health in Denmark: a survey of the attitudes and concerns of potential database participants. *Prev Vet Med*, 109(1-2):83–91, 2013b.

J Boorman. British *Culicoides*(Diptera: Ceratopogonidae): Notes on distribution and biology. *Entomol's Gaz*, 37(4):253–266, 1986.

C Foxi, G Delrio, G Falchi, MG Marche, G Satta, and L Ruiu. Role of different *Culicoides* vectors (Diptera: Ceratopogonidae) in bluetongue virus transmission and overwintering in Sardinia (Italy). *Parasit Vectors*, 9:440, 2016.

A Gerry and B Mullens. Seasonal abundance and survivorship of *Culicoides sonorensis* (Diptera: Ceratopogonidae) at a southern California dairy, with reference to potential bluetongue virus transmission and persistence. *J Med Entomol*, 37(5):675–688, 2000.

S Carpenter, M Groschup, C Garros, M Felippe-Bauer, and B Purse. *Culicoides* biting midges, arboviruses and public health in Europe. *Antivir Res*, 100(1):102–113, 2013.

A Blackwell, AJ Mordue, MR Young, and W Mordue. Bivoltinism, survival rates and reproductive characteristics of the scottish biting midge, *Culicoides impunctatus* (diptera: Ceratopogonidae) in scotland. *Bull Entomol Res*, 82(3):299–306, 1992.

A Hope. *Host location and selection by British Culicoides species associated with farms*. PhD thesis, University of Liverpool, 2013.

Dictionary. *The American Heritage Medical Dictionary*. Updated edition. Houghton Mifflin Harcourt, 2007.

S Carpenter, A Wilson, J Barber, E Veronesi, P Mellor, G Venter, and S Gubbins. Temperature dependence of the extrinsic incubation period of orbiviruses in *Culicoides* biting midges. *PloS one*, 6(11):e27987, 2011.

JT Paweska, GJ Venter, and PS Mellor. Vector competence of South African *Culicoides* species for bluetongue virus serotype 1 (BTV-1) with special reference to the effect of temperature on the rate of virus replication in *C. imicola* and *C. bolitinos*. *Med Vet Entomol*, 16(1):10–21, 2002.

EJ Wittmann, PS Mellor, and M Baylis. Effect of temperature on the transmission of orbiviruses by the biting midge, *Culicoides sonorensis*. *Med Vet Entomol*, 16(2):147–156, 2002.

PPC Mertens, JN Burroughs, A Walton, MP Wellby, H Fu, RS O'HARA, SM Brookes, and PS Mellor. Enhanced infectivity of modified bluetongue virus particles for two insect cell lines and for two *Culicoides* vector species. *Virology*, 217(2):582–593, 1996.

BA Mullens and FR Holbrook. Temperature effects on the gonotrophic cycle of *Culicoides variipennis* (diptera: Ceratopogonidae). *J Am Mosq Control Assoc*, 7(4):588–591, 1991.

BA Mullens, AC Gerry, TJ Lysyk, and ET Schmidtmann. Environmental effects on vector competence and virogenesis of bluetongue virus in *Culicoides*: interpreting laboratory data in a field context. *Vet Ital*, 40(3):160–166, 2004.

G MacDonald. The analysis of the sporozoite rate. *Trop Dis Bull*, 49(6):569–586, 1952.

S Gubbins, J Turner, M Baylis, Y Van der Stede, G van Schaik, J Abrahantes, and A Wilson. Inferences about the transmission of Schmallenberg virus within and between farms. *Prev Vet Med*, 116(4):380–390, 2014.

MH Birley and JPT Boorman. Estimating the survival and biting rates of haematophagous insects, with particular reference to the *Culicoides* obsoletus group (Diptera, Ceratopogonidae) in southern England. *J Anim Ecol*, 51(1):135–148, 1982.

L O'Connell. *Entomological aspects of the transmission of arboviral diseases by Culidoides biting midges*. PhD thesis, University of Bristol, 2002.

Met Office.

C Sanders, C Shortall, S Gubbins, L Burgin, J Gloster, R Harrington, D Reynolds, P Mellor, and S Carpenter. Influence of season and meteorological parameters on flight activity of *Culicoides* biting midges. *J Appl Ecol*, 48(6):1355–1364, 2011.

S Brand and M Keeling. The impact of temperature changes on vector-borne disease transmission: *Culicoides* midges and bluetongue virus. *J R Soc Interface*, 14(128): 20160481, 2017.

J Turner, AE Jones, AE Heath, M Wardeh, C Caminade, G Kluiters, RG Bowers, AP Morse, and M Baylis. The effect of temperature, farm density and foot-and-mouth disease restrictions on the 2007 UK bluetongue outbreak. *Sci Rep*, 9(1):112, 2019.

S Gubbins, S Carpenter, M Baylis, JLN Wood, and PS Mellor. Assessing the risk of bluetongue to UK livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *J R Soc Interface*, 5(20):363–371, 2007.

BO Nielsen. The biting midges of Lyngby Aamose (*Culicoides*: Ceratopogonidae). *Nat Jutland*, page 26, 1963.

E Klein, R Laxminarayan, DL Smith, and CA Gilligan. Economic incentives and mathematical models of disease. *Environ Dev Econ*, 12(5):707–732, 2007.

Z Balike Dieudonné. Mathematical model for the mitigation of the economic effects of the Covid-19 in the Democratic Republic of the Congo. *Plos one*, 16(5):e0250775, 2021.

RC Harris, T Sumner, GM Knight, and RG White. Systematic review of mathematical models exploring the epidemiological impact of future TB vaccines. *Hum Vaccin Immunother*, 12(11):2813–2832, 2016.

DA Ewing, BV Purse, CA Cobbold, SM Schäfer, and SM White. Uncovering mechanisms behind mosquito seasonality by integrating mathematical models and daily empirical population data: *Culex pipiens* in the UK. *Parasit Vectors*, 12:74, 2019.

M González, PM Alarcón-Elbal, J Valle-Mora, and A Goldarazena. Comparison of different light sources for trapping *Culicoides* biting midges, mosquitoes and other dipterans. *Vet Parasitol*, 226:44–49, 2016b.

E Viennet, C Garros, I Rakotoarivony, X Allene, L Gardès, J Lhoir, I Fuentes, R Venail, D Crochet, R Lancelot, M Riou, C Moulia, T Baldet, and T Balenghien. Host-seeking activity of bluetongue virus vectors: endo/exophagy and circadian rhythm of *Culicoides* in Western Europe. *PloS one*, 7(10):e48120, 2012.

KR Snow. *Mosquitoes*. Number 14. Pelagic Publishing, 1990.

JA Campbell and EC Pelham-Clinton. X.—A Taxonomic Review of the British Species of *Culicoides* Latreille (Diptera, Ceratopogonidæ). *Proc R Soc Edinb B*, 67(3):181–302, 1960.

C Simon, F Frati, A Beckenbach, B Crespi, H Liu, and P Flook. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am*, 87(6):651–701, 1994.

K Lehmann, D Werner, B Hoffmann, and H Kampen. PCR identification of culicoid biting midges (Diptera, Ceratopogonidae) of the Obsoletus complex including putative vectors of bluetongue and Schmallenberg viruses. *Parasit Vectors*, 5:213, 2012.

S Carpenter, HL Lunt, D Arav, GJ Venter, and PS Mellor. Oral susceptibility to bluetongue virus of *Culicoides* (Diptera: Ceratopogonidae) from the United Kingdom. *J Med Entomol*, 43(1):73–78, 2006.

R Meiswinkel, P Van Rijn, P Leijs, and M Goffredo. Potential new *Culicoides* vector of bluetongue virus in northern Europe. *Vet Rec*, 161(16):564–565, 2007.

PS Mellor. The replication of bluetongue virus in *Culicoides* vectors. *Curr Top Microbiol Immunol*, 162:143–161, 1990.

PS Mellor, J Boned, C Hamblin, and S Graham. Isolations of African horse sickness virus from vector insects made during the 1988 epizootic in Spain. *Epidemiol Infect*, 105(2):447–454, 1990.

N De Regge, I Deblauwe, R De Deken, P Vantieghem, M Madder, D Geysen, F Smeets, B Losson, T Van den Berg, and AB Cay. Detection of Schmallenberg virus in different *Culicoides* spp. by real-time RT-PCR. *Transbound Emerg Dis*, 59(6):471–475, 2012.

T Balenghien, N Pagès, M Goffredo, S Carpenter, D Augot, E Jacquier, S Talavera, F Monaco, J Depaquit, C Grillet, J Pujols, G Satta, M Kasbari, M-L Setier-Rio, F Izzo, C Alkan, J-C Delécolle, M Quaglia, R Charrel, A Polci, E Bréard, V Federici, C Cêtre-Sossah, and C Garros. The emergence of Schmallenberg virus across *Culicoides* communities and ecosystems in Europe. *Prev Vet Med*, 116(4):360–369, 2014.

GE Chapman, K Sherlock, JC Hesson, MSC Blagrove, GJ Lycett, D Archer, T Solomon, and M Baylis. Laboratory transmission potential of British mosquitoes for equine arboviruses. *Parasit Vectors*, 13:413, 2020.

European Centre for Disease Prevention and Control. *Aedes aegypti - current known distribution: March 2021*. URL `https://www.ecdc.europa.eu/en/publications-data/aedes-aegypti-current-known-distribution-march-2021`. Accessed: 2021-09-29.

T Dallimore, T Hunter, JM Medlock, AGC Vaux, RE Harbach, and C Strode. Discovery of a single male *Aedes aegypti* (L.) in Merseyside, England. *Parasit Vectors*, 10:309, 2017.

K Roosa and G Chowell. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theor Biol Med Model*, 16(1):1, 2019.

C Okaïs, S Roche, ML Kürzinger, B Riche, H Bricout, T Derrough, F Simondon, and R Ecochard. Methodology of the sensitivity analysis used for modeling an infectious disease. *Vaccine*, 28(51):8132–8140, 2010.

MD McKay, RJ Beckman, and WJ Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

SM Blower and H Dowlatabadi. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int Stat Rev*, pages 229–243, 1994. doi: 10.2307/1403510.

A Saltelli, M Ratto, T Andres, F Campolongo, J Cariboni, D Gatelli, M Saisana, and S Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

J Wu, R Dhingra, M Gambhir, and JV Remais. Sensitivity analysis of infectious disease models: methods, advances and their application. *J R Soc Interface*, 10(86):20121018, 2013.

IM Sobol'. On sensitivity estimation for nonlinear mathematical models. *Matem Mod*, 2(1):112–118, 1990.

IM Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simul*, 55(1-3):271–280, 2001.

DA Rand. Mapping global sensitivity of cellular network dynamics: sensitivity heat maps and a global summation law. *J R Soc Interface*, 5(suppl_1):S59–S69, 2008.

sassyMATLAB Toolbox. sassyMATLAB Toolbox, 2013. https://github.com/pkrusche/sassy-helpers.

MD Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

F Campolongo, J Cariboni, and A Saltelli. An effective screening design for sensitivity analysis of large models. *Environ Model Softw*, 22(10):1509–1518, 2007.

A Saltelli, S Tarantola, F Campolongo, and M Ratto. *Sensitivity analysis in practice: a guide to assessing scientific models*. Wiley Online Library, 2004.

M Kendall and A Stewart. *The Advanced Theory of Statistics*. Hafner Publishing Company, New York, 1979.

M Domijan and DA Rand. Balance equations can buffer noisy and sustained environmental perturbations of circadian clocks. *Interface Focus*, 1(1):177–186, 2010.

C Viboud, L Simonsen, and G Chowell. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics*, 15:27–37, 2016.

G Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infect Dis Model*, 2 (3):379–398, 2017.

FJ Richards. A flexible growth function for empirical use. *J Exp Bot*, 10(2):290–301, 1959.

L Dinh, G Chowell, K Mizumoto, and H Nishiura. Estimating the subcritical transmissibility of the Zika outbreak in the State of Florida, USA, 2016. *Theor Bio Med Model*, 13(1):20, 2016.

DE Parker, TP Legg, and CK Folland. A new daily central England temperature series, 1772–1991. *Int J Climatol*, 12(4):317–342, 1992.

AE Jones, J Turner, C Caminade, AE Heath, M Wardeh, G Kluiters, PJ Diggle, AP Morse, and M Baylis. Bluetongue risk under future climates. *Nat Clim Chang*, 9(2):153–157, 2019.

JA Backer and G Nodelijk. Transmission and control of African horse sickness in The Netherlands: a model analysis. *PLoS One*, 6(8):e23066, 2011. doi: 10.1371/journal.pone.0023066.

APPENDIX A

Surveillance of UK vector species

A.1 Methods

Vector collection and morphological identification Vectors were caught in two BG-Sentinel Traps (Biogents AG, Regensburg, Germany) with green LED lights near the entrance [González et al., 2016b] at host premises during the summer of 2020. Trapping was carried out at both an equine and human premises located on The Wirral Peninsula, Merseyside, UK. Traps were emptied daily at between 5 hours after sunrise and 5 hours before sunset [Viennet et al., 2012]. Insects were then frozen for at least 4 hours before larger insects (including mosquitoes) were frozen and smaller insects (including *Culicoides* were stored in 70% ethanol for further analysis. Mosquitoes [Snow, 1990] and *Culicoides* [Campbell and Pelham-Clinton, 1960] were classified morphologically to species level. The insects were then sexed [Snow, 1990] and parity of *Culicoides* was assessed (Section 1.6).

Molecular identification of vectors For suspected vector samples that were not identifiable by morphology, DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). Molecular analyses of insects were made using the primers C1-J-1718 and C1-N-2191 (Table A.1), which amplify 450 bp of the COI gene. One insect was identified as a member of the *Culicoides* Obsoletus group, but disfigurement prevented a full species identification. Molecular analysis for this insect was made using the dew-COI-fwd and obs-COI-fwd forward primers paired with C1-N-2191. A polymerase chain reaction (PCR) of 50 μL was made, containing 1 μL of each primer, 25 μL MyFi Reaction Buffer (Bioline, London, UK), 2 μL of the DNA template and 21 μL of nuclease-free water. The PCR consisted of an initial heating phase at 95°C for 1 min, followed by 30 cycles of denaturation (95°C for 30s), annealing (50°C for 30s), elongation (72°C for 30s), and a final elongation step (72°C for 1 min). Positive and negative controls were used with every PCR. Amplifications were confirmed by loading the PCR products were loaded onto a 2% w/v agarose in TAE buffer gel stained with Nancy-520 (Sigma-Aldrich, St. Louis, Missouri, United States) and then scanned to confirm amplifications. The unidentifiable sample from the *Culicoides* Obselotus group was identified by the location of the positive band. PCR products displaying positive bands on the gel were purified using a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and submitted to Source Bioscience, Nottingham UK for Sanger sequencing using the C1-N-2191 primer. Successful matches for the sequences were obtained using the standard nucleotide Basic Local Alignment Search Tool (BLAST) available on the National Center for Biotechnology Information (NCBI) website.

Primer	5′-3′ sequence	Ref.
C1-J-1718 (F)	GGAGGATTTGGAAATTGATTAGTTCC	[Simon et al., 1994]
C1-N-2191 (R)	CCCGGTAAAATTAAAATATAAACTTC	[Simon et al., 1994]
dew-COI-fwd (F)	CGCCCGACATAGCATTCCCT	[Lehmann et al., 2012]
obs-COI-fwd (F)	CAGGAGCTTCTGTAGATTTGGCT	[Lehmann et al., 2012]

Table A.1: Primer sequences used in the molecular analysis for vector species identification. Parentheses after the primer name indicate forward (F) or reverse (R) direction.

Culicoides species	Nuliparus	Blood-fed	Gravid	Parus
C. dewulfi	0	4	2	3
C. halophilus	1	0	0	0
C. impunctatus	2	0	0	5
C. pulicaris	0	0	0	2
C. obsoletus	0	0	0	1
C. nubeculosus	1	0	0	0

Table A.2: Parity of caught *Culicoides* by species.

A.2 Results

Overall six species of *Culicoides* and two species of mosquitoes were caught. *Culex pipiens* were caught in both equine and human premises, whereas all other species were only caught at the equine premises. A summary of the *Culicoides* and Mosquitoes caught it shown in Figure A.1.

Culicoides The *Culicoides* species included *C. dewulfi* (9), *C. halophilus* (1), *C. impunctatus* (7), *C. pulicaris* (2), *C. obsoletus* (1) and *C. nubeculosus* (1). The parity by species is given in Table A.2. *C. obsoletus*, *C. pulicaris* and *C. nubeculosus* are confirmed and *C. dewulfi* is a suspected vector of BTV [Carpenter et al., 2006, Meiswinkel et al., 2007, Mellor, 1990]. The *C.obsoletus* group and *C. pulicaris* have also been implicated in the transmission of AHSV [Mellor et al., 1990]. *C. obsoletus*, C. dewulfi, *C. pulicaris* and *C. nubeculosus* have been implicated in SBV [De Regge et al., 2012, Balenghien et al., 2014].

Mosquitoes The mosquito species included *Culex pipiens*, vector of West Nile virus and Japanese encephalitis virus [Chapman et al., 2020], and *Aedes aegypti*, vector of dengue, Zika, chikungunya and yellow fever viruses. Although *Aedes aegypti* is officially absent from this region [Centre for Disease Prevention and Control] there has been a male previously caught in the region [Dallimore et al., 2017]. *Culex pipiens* included 19 females (of which only one was caught at the human premises) and 4 males (of which 2 were caught at both the equine and human premises).

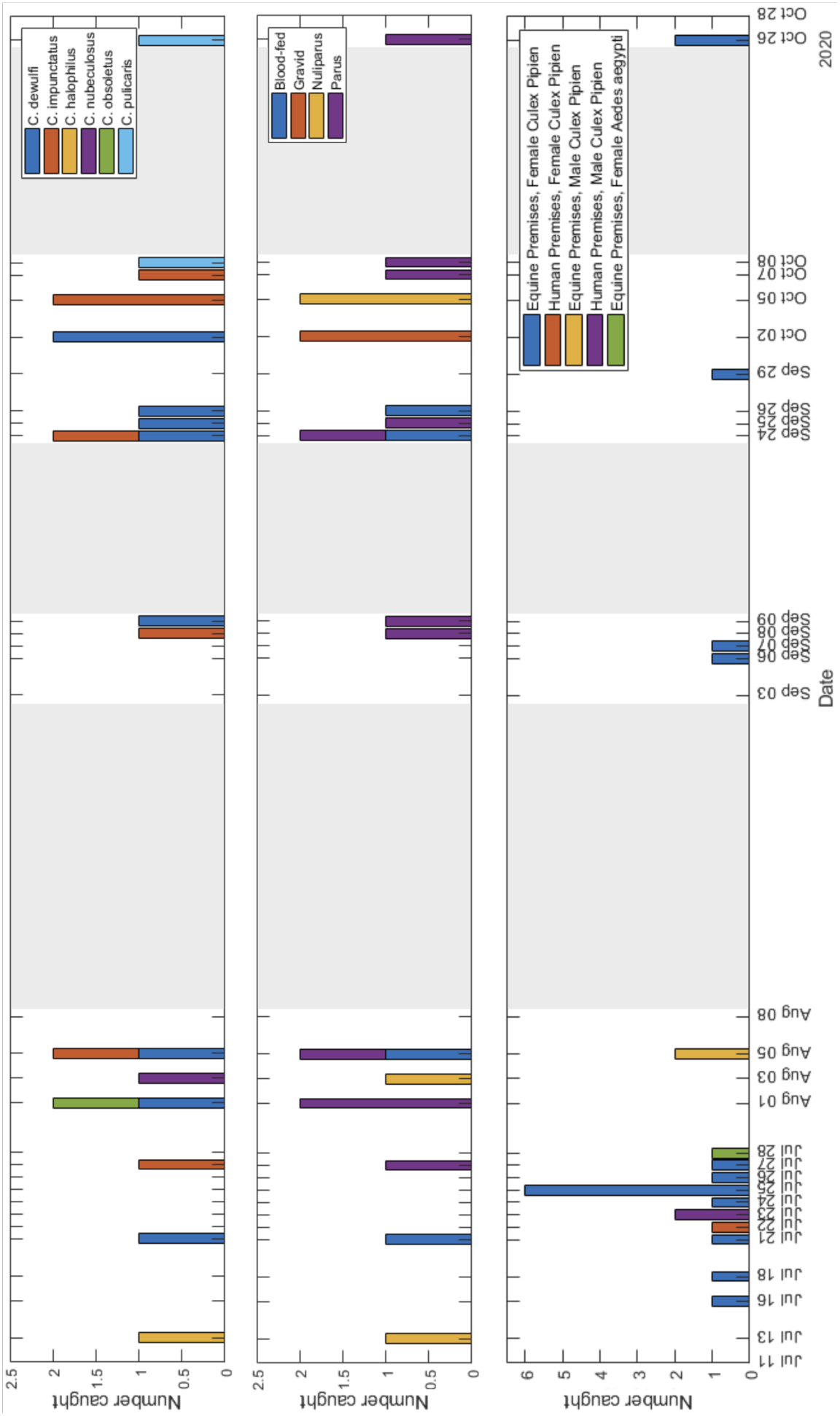

Figure A.1: Trapped *Culicoides*, by species (top) and parity (middle), and mosquitoes by species, gender and location (bottom). Shaded areas represent time periods with no trapping.

APPENDIX B

Assessing parameter identifiability

B.1 Introduction

Mathematical models are increasingly used to guide policy makers in decisions regarding interventions. However, in order for this to be successful, we need to be able to reliably estimate model parameters and their uncertainty. In some models, multiple sets of parameters can yield similar simulation results; these are non-identifiable. Non-identifiability can be caused by the data-set (number of observations, observation error, spatial-temporal resolution), known as practical identifiability, or the structure of the model, known as structural identifiability.

We used a parametric bootstrap approach described in Roosa and Chowell [2019] to assess the identifiability of parameters. Using known parameters, the models were simulated to quantify the uncertainty and identifiability of the parameters. First we repeated the analysis of the basic SEIR model, given as performed by Roosa and Chowell [2019] on the 1918 influenza pandemic in San Francisco (California), to demonstrate the methods used. We then adapted this method to consider a common issue within disease data; under-reporting, and observed how this affects the outcomes.

B.2 A parametric bootstrap method

To assess parameter identifiability of a model, we first fit the deterministic solution time series data (F) using the *lsqcurvefit* function in Matlab to find the best fit using nonlinear least squares estimation. The parameters used to obtain F are used as the prior. Then S data-sets are created from F with Poisson error structure for the daily incidences ($\dot{C}(t)$). Here for each time-point a new incidence value is generated using a Poisson random variable with mean $\dot{C}(t)$. Then we select S parameter sets within a realistic parameter space using Latin hypercube sampling (LHS) and use these as our priors for the parameter estimation. This differs from the original method described by Roosa and Chowell [2019], where the priors of the parameters varied according to a uniform distribution in the range of ± 0.1 around the true value, as LHS allows us to explore the whole parameter space. Each of these S parameter sets is then fitted to one S data-set using nonlinear least squares approximation i.e. searching for the parameter set that minimises the sum of squared differences between the Poisson error structured data-set and the model solution. We begin by estimating each parameter individually, while the rest of the parameters are fixed, and then increase the number of parameters being jointly estimated. This allows us to observe which parameters can be estimated

together with a reasonable degree of accuracy and construct confidence intervals for each estimated parameter. We also calculate R_0 for each estimated parameter set and observe its identifiability.

If a parameter is identifiable from the available data its confidence interval lies in a finite range of values. Here we calculate the 5–95% confidence interval for parameters. The smaller the confidence interval the more identifiable the parameter is. For each of the S repeats the mean squared error (MSE) is calculated for each parameter and R_0 given by

$$MSE = \frac{1}{S}\sum_{i=1}^{S}(\theta - \hat{\theta}_i)^2, \tag{B.1}$$

where θ represents the true parameter value and $\hat{\theta}_i$ represents the estimated parameter from the i^{th} prior and data-set. A smaller MSE would suggest a more identifiable parameter.

SEIR model (Human influenza)

As an example, we will demonstrate this method on a SEIR model of human influenza that consists of four parameters (transmission rate, length of the latent period, length of the infectious period and the population size) and 4four compartments (susceptible, latently infected, infectious and recovered). The population size, N, is considered to be constant ($N = S + E + I + R = 500,000$) and natural or disease-caused death is not considered. The system of ordinary differential equations (ODEs) is therefore:

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta}{N}SI, \\
\frac{dE}{dt} &= \frac{\beta}{N}SI - \kappa E, \\
\frac{dI}{dt} &= \kappa E - \gamma I, \\
\frac{dR}{dt} &= \gamma I, \\
\frac{dC}{dt} &= \kappa E,
\end{aligned} \tag{B.2}$$

where $\beta = 0.56$ is the transmission rate, $1/\kappa = 1.9$ is the latent period, $1/\gamma = 4.1$ is the infectious period and the variable C tracks the cumulative number of infectious individuals since the start of the outbreak. Therefore the daily incidence curve is given by $\dot{C}$. Some basic analysis gives $R_0 = \beta/\gamma = 2.3$ for this model. Since the whole realistic parameter space is not given for β, γ and κ we will use a uniform distribution of ± 0.1, similarly to the original study.

Under-reporting

Under-reporting of cases will add noise to the data, which may effect the identifiability of the parameters. Under-reporting may occur when a disease does not have many visible clinical signs and therefore goes undetected, for example, Schmallenberg virus (SBV) and West Nile virus (WNV). Here we initially generated data-sets using the true parameter values and no under-reporting; the time-points for the S Poisson distributed parameter sets are now generated as the product of a Poisson random variable with

mean $\dot{C}(t)$ and 1-r, where r is the fraction of cases not reported.

Under-reporting may be incorporated into the model by modifying the ODE for the cumulative number of cases (C) in Equations B.2 to be

$$\frac{dC}{dt} = \hat{r}kE, \tag{B.3}$$

where $\hat{r}$ is the fraction of cases not reported. However our prior for $\hat{r}$ may influence the output of the model. The number of parameters that need to be jointly estimated also increases.

Vector-borne disease model (African horse sickness virus)

We also use this method to examine the parameter identifiablity of the AHSV ODE model from Chapter 3.

B.3 Results

B.3.1 SEIR model (human influenza)

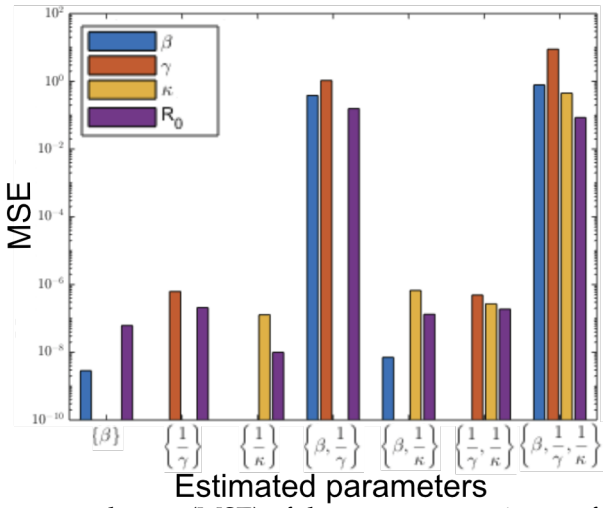

Figure B.1: Mean squared error (MSE) of the parameter estimates for each estimated parameter set (S=200) for the SEIR model.

In the model of human influenza, all parameters are identifiable when only identifying one parameter as they all have MSEs $< 10^{-6}$ and small confidence intervals (Figures B.1 and B.2). However, when trying to jointly identify two parameters β and γ are not jointly identifiable, with MSEs $> 10^{-1}$ and large confidence intervals spanning out of the original assumed distributions. The medians of β and κ do give a more accurate estimation than the means. The distribution appears to be skewed with most of the parameters close to the true value but also some very poor estimates. Whereas, β and κ and κ and γ can be jointly identified with similar confidence intervals and MSEs as when only one parameter was identified. When attempting to jointly identify β, κ and γ, large MSEs are observed, similar to when jointly identifying β and γ, however the confidence intervals are small and close to the true value. This suggests that the model fits quite well in most circumstances.

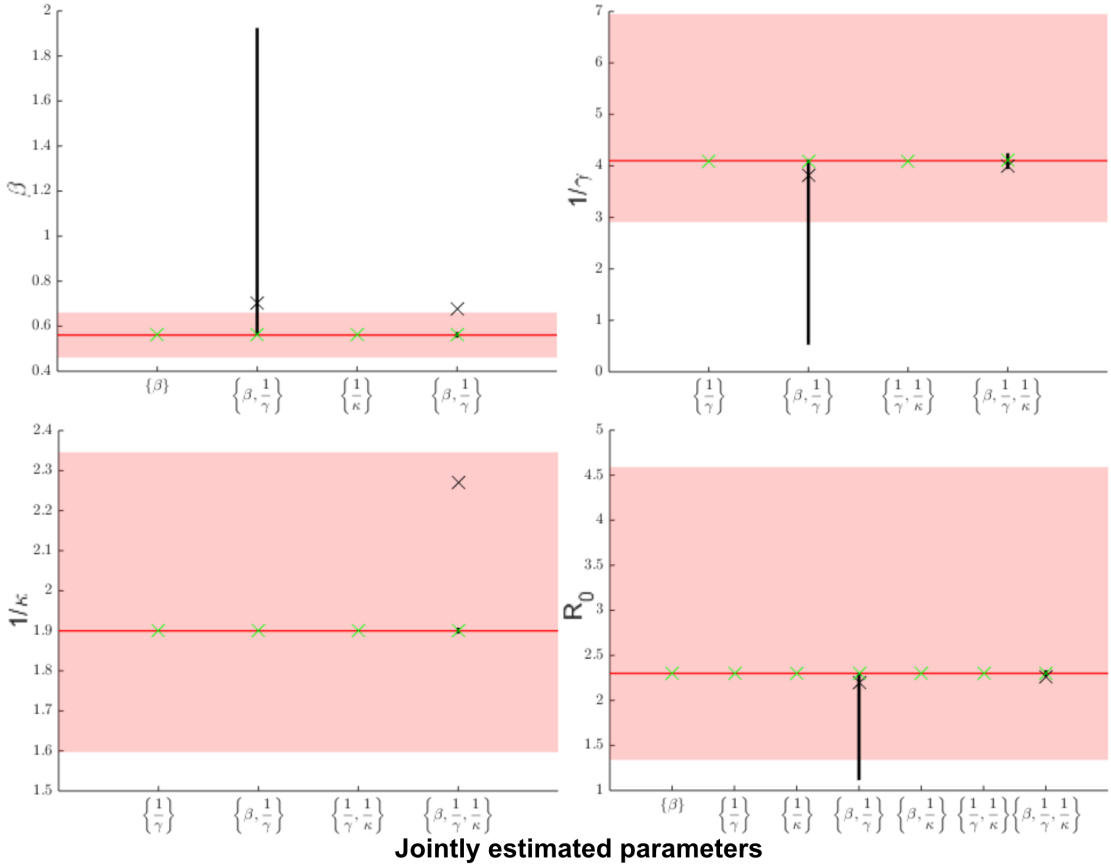

Figure B.2: Results from the bootstrap method for assessing identifiability for the simple SEIR model. The parameter estimates $5-95\%$ confidence intervals (black vertical lines), mean (black x) and median (green x) for (a) β, (b) $1/\gamma$, (c) $1/\kappa$, and (d) R_0, for each estimated parameter set (S=200). The true parameter value is represented by the filled red horizontal line and the dashed red horizontal lines represent the upper and lower limits of the parameters used as priors.

B.3.2 Effects of under-reporting

Figure B.3 shows how the identifiability of parameters changes as the fraction of cases not reported, r, increases between 0 to 0.5. We observe that as r increases the MSE increases for all estimates. Also, the sets of jointly estimated parameters that gave the largest MSE values and confidence intervals in Section B.3.1 ($\{\beta, 1/\gamma\}$ and $\{\beta, 1/\gamma, 1/\kappa\}$) produce unsmooth lines, whereas the other parameter sets produce smooth lines. From now on the sets of jointly estimated parameters which produce smooth line will be referred to as the stable parameter sets. We observe that these stable parameter sets tend to increasingly underestimate the latent period ($1/\gamma$), the infectious period ($1/\kappa$) and the basic reproduction number (R_0). The transmission rate (β) is also underestimated for stable parameter sets, however to a lesser extent than the other parameters. For low levels of under-reporting parameter estimates are still fairly accurate, however as r increases the accuracy of estimations quickly decreases.

B.3.3 Vector-borne disease model (African horse sickness virus)

Figure B.4 shows results for the parameter estimates for the AHSV ODE model (Chapter 3) when one parameter is estimated and all other parameters are fixed for each parameter. There are no results for the proportion of hosts that die (m_H) as this parameter could not be estimated using the *lsqcurvefit* function. We observe that we obtain reasonable distributions for the host latent period ($1/\epsilon$), infectious period (T^1_{inf}), vector life-span ($1/\mu_V$), transmission probability from vector to host (p_V) and the initial number of hosts (N_H). The initial number of hosts is likely to be known when modelling the spread of disease at an equine premises. However, if a vector-borne disease had a reservoir host then this value may not be known. An example of this is WNV since the exact number of birds, which are the reservoir hosts, may not be known. We observe a uniform distribution for the vector to host ratio (ρ); this is uninformative and suggests that we are not able to estimate this parameter. The distributions for the blood feeding interval ($1/a$), EIP ($1/\upsilon$) and transmission probability from host to vector (p_H) are bi-modal. This suggests that the model is not structurally identifiable for these parameters. One of the peaks is at the true value; therefore, if the other peak is unrealistic, restricting the upper and lower limits of estimates such that the unrealistic peak is outside these limits could improve results. However, this may not always be the case. We observe that the mean squared error is largest for ρ, which is not identifiable, however it does not reflect the non-identifiablity of $1/a$, $1/\upsilon$ or p_H.

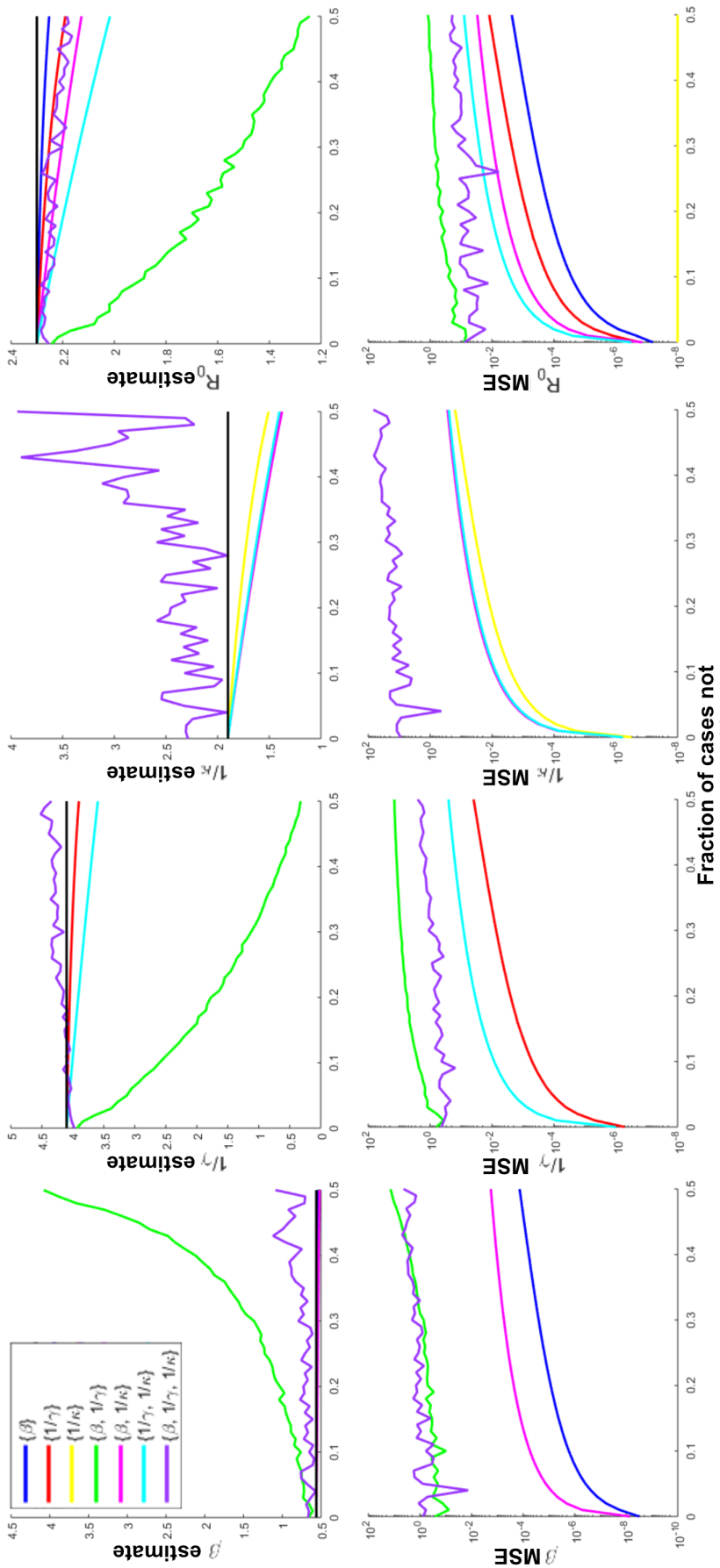

Figure B.3: The parameter estimates and mean squared errors for the parameters and R_0 in the deterministic SEIR model for different fractions of under-reporting.

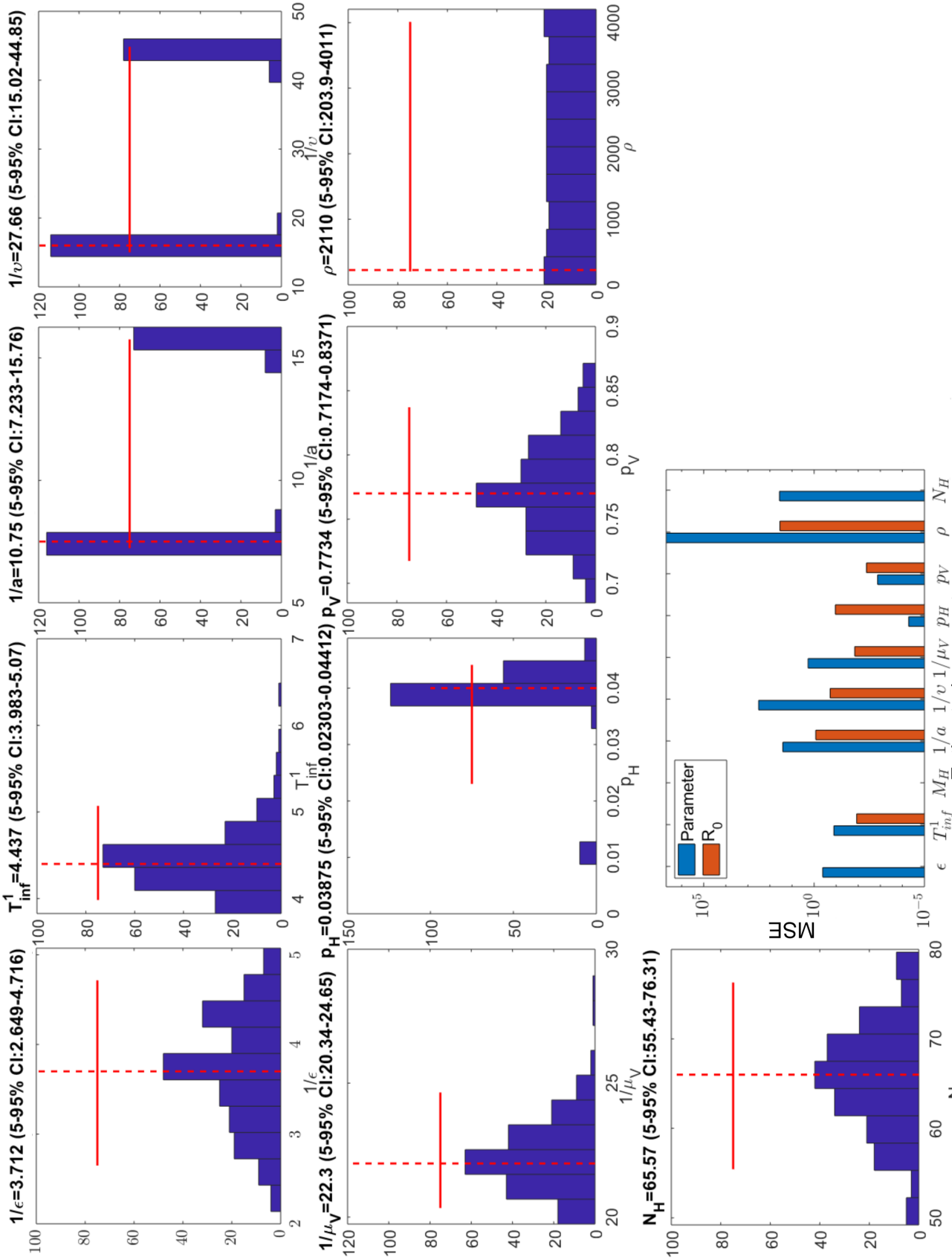

Figure B.4: Results for the AHSV ODE model for singularly estimated parameters. The histograms represent the $s = 200$ estimates where the vertical red dashed line represents the true values and the horizontal red lines represent the confidence intervals. The bar chart shows the mean MSEs of each parameter and corresponding R_0 estimates.

APPENDIX C

Sensitivity Analysis Methods

C.1 Introduction

Simple models can be solved analytically, however as models increase in complexity numerical analysis is needed to understand their behaviour. These techniques include uncertainty and sensitivity analysis. As we increase the number of parameters jointly estimated the uncertainty of our estimations increases. Uncertainty analysis may therefore be used to explore variability of outcome variables due to the uncertainty of input parameters. Sensitivity analysis quantifies the influence input parameters have on the output variables, and therefore highlights which parameters are the most important. If a model is extremely sensitive to a small change in a parameter that is imprecise it may bias the prediction for the transmissibility of the disease and/or the impact of control measures [Okaïs et al., 2010].

The key qualities of a sensitivity analysis method can be considered to be:

- The ability use on stochastic models - so all models can be analysed using the same technique, allowing for comparison between model types.
- The ability to measure the variation in sensitivity over time - to access the appropriate timing of control measures and the influence of parameters on the duration/peak incidence of the outbreak.
- Shows the direction of sensitivity - to access the appropriate timing of control measures and the influence of parameters on the duration/peak incidence of the outbreak.

A challenge in sensitivity analysis is how parameters are sampled; different methods of sampling come with different computational costs. Parameters could be sampled using a full factorial sampling design, which uses all possible values of each parameter and each possible combination of parameters. An advantage of this method is that the whole parameter space is explored, however it is very time consuming for complex models. Another method could be to just vary one parameter at a time over a specified range. However, although it is a lot faster, this only allows us to explore a small amount of the possible parameter space and the precision of the other parameters could affect the results. A more sophisticated statistical analysis technique is Latin Hypercube Sampling (LHS), which allows us to vary all input parameters simultaneously [McKay et al., 1979]. This method therefore allows us to explore a large volume

of the parameter space while only using each value of each parameter once, making it very efficient. By calculating the partial rank correlation coefficients (PRCC) for each input parameter and each outcome variable we can perform a sensitivity analysis using the LHS scheme [Blower and Dowlatabadi, 1994].

There are multiple possible ways of performing a sensitivity analysis. One of the simplest ways to visually examine the effect a parameter has on the model output as a scatter plot. If the model output is sensitive to the parameter we would expect to see a correlation between the parameter and output. Parameters are usually sampled using a Monte Carlo method. A disadvantage of this method is that we can only test the sensitivity of one parameter at a time [Saltelli et al., 2008, Wu et al., 2013]. The Sobol' method is able to attribute the variance of a model output to each parameter and show the interaction effects on sensitivity between parameters, which the Morris and LHS-PRCC methods cannot. However, it has a very high computation cost and cannot show the sensitivity over time or show the direction of sensitivity [Sobol', 1990, 2001]. As the Morris and LHS-PRCC methods are able to do these, it was decided they would be more suitable. Of these methods, the Morris method is the easiest to implement; the LHS-PRCC and Sobol' methods are equally complex and difficult to implement [Wu et al., 2013].

Recently, new graphical techniques have been developed: the sensitivity heat map (SHM) method and the parameter sensitivity spectrum (PSS) [Rand, 2008]. The PSS shows the sensitivity of a single parameter to multiple output variables, whereas the SHM shows the effects of multiple parameters on a single output variable. Heat maps are generated such that we can assess how the sensitivity of parameters change over a time course. This could provide useful information on the most appropriate timing for control strategies. These methods are relatively computationally efficient, although they do not show the direction of sensitivity. Although the concept behind the two methods is complicated, they are fairly easily executed using the SASSY Matlab package [sassyMATLAB Toolbox]. This also introduces limitations as the toolbox only allows scalar or rate rules and is limited to ordinary differential equation (ODE) models. This raises concerns as to how well this method will cope with temperature being incorporated into a model.

Which of the desirably qualities of a sensitivity analysis each method satisfies is summarised in Table C.1.

Method	Stochastic models	Variation over time	Direction of sensitivity
Morris	yes	yes	no
LHS-PRCC	yes	no	yes
SHM	no	yes	no
Scatter plots	yes	no	yes
Sobol'	yes	no	no

Table C.1: Desired qualities of sensitivity analysis possessed by the methods considered [Wu et al., 2013].

C.2 Methods considered

The Morris method (elementary effects method) The Morris method considers the ratio of the change in an output variable, $y_j \in Y$, to the change in a parameter, $x_i \in X$, where X and Y are the model input parameters and outputs, respectively [Morris, 1991]. The parameter spaces, for each of the k input parameters, are divided into p sections which the parameters are sampled from using repeated random sampling. The elementary effect of x_i on y_j can be expressed as

$$EE_i(X) = \frac{y_j(x_1, x_2, \ldots, x_i + \Delta, x_{i+1}, \ldots, x_k) - y_j(X)}{\Delta}, \tag{C.1}$$

where the K input parameters are given as a scaled vector $X \in [0,1]^k$ and Δ is a value in the set $\{1/(p-1), \ldots, 1 - 1/(p-1)\}$. This is repeated for all the selected parameter samples, giving a distribution, denoted F_i. The mean of F_i and $|F_i|$, denoted μ and μ^*, respectively, as well as the standard deviation of F_i, σ are the sensitivity measures used [Campolongo et al., 2007, Wu et al., 2013]. If the value of μ is large it is a sign that the model output is sensitive to this parameter. If σ is large it suggests that there is a non-linear relationship between the parameter and output or that this parameter interacts with other parameters [Saltelli et al., 2004, Wu et al., 2013]. The absolute elementary effect, μ^* is used to check that a small μ is not produced due to F_i containing both positive and negative values. In this case elementary effects may cancel each other out, producing a small μ even though the input parameter is sensitive.

Partial rank correlation coefficient method (LHS-PRCC method) The LHS-PRCC technique requires us to first define probability distribution functions for each of the parameters we wish to explore using existing knowledge. We can then use the method described in Blower and Dowlatabadi [1994]. We start by generating a matrix of N values for each of the K parameters we wish to test the sensitivity of sampled using Latin hypercube sampling (LHS), giving a N by K matrix. The lower limit for the number of samples, N, is $K + 1$; since the LHS design uses sampling without replacement if only K draws are made the K^{th} draw would be predetermined. This value is also affected by the significance level we desire for the PRCC.

We then calculate the Q output variables for each of these N parameter sets, giving a N by Q matrix. For each of the input parameters ($k \in K$) and the output parameters ($q \in Q$) we rank the magnitude of all the samples (1 corresponds to the smallest sample and N to the largest sample) defined as the set $(r_{1,i}, r_{2,i}, \ldots, r_{K,i}, R_{1,i}, R_{2,i}, \ldots, R_{Q,i})$ where i is the run number, and r and R correspond to the ranks of the parameters and outputs, respectively. The average rank is therefore given by $\mu = (N+1)/2$. It is now possible to define a $K + Q$ by $K + Q$ matrix C with elements calculated as:

$$c_{i,j} = \frac{\sum_{n=1}^{N}(r_{in} - \mu)(r_{jn} - \mu)}{\sqrt{\sum_{m=1}^{N}(r_{im} - \mu)^2 \sum_{n=1}^{N}(r_{jn} - \mu)^2}}, \; i, j = 1, 2, \ldots, K, \tag{C.2}$$

and

$$c_{i,K+q} = \frac{\sum_{n=1}^{N}(r_{in} - \mu)(R_{q,i} - \mu)}{\sqrt{\sum_{m=1}^{N}(r_{im} - \mu)^2 \sum_{n=1}^{N}(R_{q,i} - \mu)^2}}, \; i, q = 1, 2, \ldots, Q, . \tag{C.3}$$

The leading diagonal of C should therefore all be ones. The matrix B, with elements b_{ij}, is defined as the inverse of C [Kendall and Stewart, 1979]. The PRCC between the i^{th} input parameters and the y^{th} output variable is given by

$$\gamma_{iy} = \frac{-b_{i,K+1}}{\sqrt{b_{ii}b_{K+1,K+1}}}. \tag{C.4}$$

A Student's T test with $N-2+\zeta-1$ degrees of freedom, where ζ is the number of parameters jointly sampled using LHS, can be used to test the significance of any nonzero γ_{iy} values, with t-values calculated as

$$t_{iy} = \gamma_{iy}\sqrt{\frac{N-2}{1-\gamma_{iy}}}. \tag{C.5}$$

If the PRCC is close to 0 the parameter is weakly (or not at all) correlated with the output parameter. As the PRCC approaches 1 or -1 it represents a positive or negative correlation, respectively, with 1 and -1 reflecting a perfect correlation.

Sensitivity heat maps The SHM and PSS methods consider systems of n ODEs with k input parameters given as:

$$\frac{dY}{dt} = f(t, Y_0, X), \tag{C.6}$$

where t is time, $y_j \in Y$ are the vector state variables, Y_0 are the initial values of these state variables and $x_i \in X$ are the parameters in the model. There will be a single or set of solutions $Y = g(t, X)$ for $0 \leq t \leq T$. A change in parameters (δX) yields a change in the solution (δg) given as $\delta g = M\delta X + O(||\delta X||^2)$, where M is the linear transformation from parameter space $\mathbb{R}^2$ to a Hilbert space H [Domijan and Rand, 2010]. Then values U, W and $\{\sigma_i\}$, which express sensitivity, can be derived through singular value decomposition of a time normalised matrix composed of the partial derivatives $\delta g_i/\delta x_i$ at each simulation time point [Rand, 2008]. Rand [2008] showed that the PSS for a parameter and SHM for a state variable ($g_i(t)$) can the be expressed by $\delta g(t)/\delta k_j = \Sigma_i \sigma_i W_{ij} U_i(t)$ and $\sigma_i(max_j|W_{ij})U_i(t)$, respectively.

C.3 Application of new method

In order to develop a method which shows each of the key qualities desired (Table C.1) we propose that the LHS-PRCC is used with each simulation time-step as an output. This would allow all three of these characteristics to be shown and could produce heat maps for models that are not based on ODEs. Although this method will be more computationally expensive than the original method, for the number of time-steps in most epidemic models it is still reasonable.

We perform the original LHS-PRCC method on R_0, where $R_0 = \beta/\mu$, and the heatmap LHS-PRCC method on the case incidence for the simple SEIR model given as:

$$\frac{dS}{dt} = -\beta SI, \frac{dE}{dt} \qquad = \beta SI - \sigma E, \frac{dI}{dt} = \sigma E - \mu I, \frac{dR}{dt} \qquad = \mu I. \tag{C.7}$$

This methods is also used in a vector-borne disease model (Chapter 3).We perform the calculation of the PRCC described in Section C.2 100 times for 100 sets of parameters

each time, the median PRCC and its variation are examined using a box-plot. For the heat map method the PRCC is calculated only once per time-step and is performed on 100 time for 100 parameters sets; the number value plotted is the mean PRCC of the outputs. The heat maps will vary from white to red for 0 to 1 and from white to blue from 0 to -1. Therefore the more positively correlated a parameter is the more red it will be, and the more negatively correlated a parameter is the more blue it will be. If a result was not found to be significant through the Student's t-test ($p > 0.05$) it was coloured white. Matlab was used for calculations and to produce figures, LHS was performed using the *lhsdesign* function. The parameters were all selected for a uniform distribution (upper and lower bounds shown in Table C.2). These bounds were calculated as ± 0.1 on the rates β, σ and μ, similarly to [Roosa and Chowell, 2019].

Para.	Description	True value	Lower bound	Upper bound
β	Transmission rate (per day)	0.56	0.46	0.66
$1/\sigma$	Mean latent period (days)	1.9	1.59	2.33
$1/\mu$	Mean infectious period (days)	4.1	2.94	7.14

Table C.2: Lower and upper bounds used in LHS for each parameter (para) in the simple SEIR model [Roosa and Chowell, 2019].

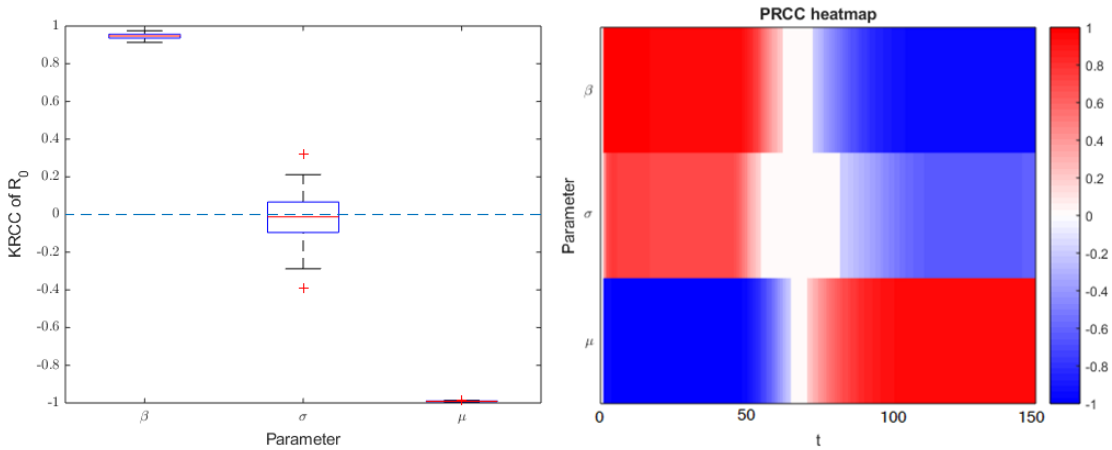

Figure C.1: Results from sensitivity analysis of the simple SEIR model. Left: PRCC of parameters and R0. A Student's t-test was performed and σ was not found to significantly influence R_0 ($p > 0.05$). Right: The mean PRCC of parameters on the time-series of incidences.

The original PRCC method with R_0 as the output (Figure C.1) shows that β is positively correlated with R_0 (mean = 0.95), this means that as β increases we would expect R_0 to increase. Since μ is negatively correlated with R_0 (median = -0.99) the larger μ is, the smaller we would expect R_0 to be. Since the infectious period is given by $1/\mu$, longer infectious periods give a larger R_0. Since σ is not significant ($p = 0.92$) the latent period ($1/\sigma$) is not significant, this is expected because σ is not used to calculate R_0.

The heat map PRCC (Figure C.1) shows that β is positively correlated with incidences until $t \simeq 70$ then it becomes negatively correlated. This is because increasing β shortens the duration of the outbreak, therefore more cases are occurring earlier. However,

overall there are more cases when β is increased. This is not clearly visible from this heat map result. Since μ is negatively correlated at the beginning of the time-course and positively correlated at the end we conclude that a longer infectious period also shortens the duration of the outbreak. Again, we cannot see that the lengthened infectious period also increases the total number infected. We also see that the incubation rate (σ) also affects daily incidence in a similar way to β. Therefore, a longer latent period causes the outbreak to occur later. This is not observable using the R_0 PRCC method. However, since the red/blue on the heat map is not as bright for σ we observe this influence is weaker than for the other parameters.

As the heat map cannot show the influence parameters have on the total number of infected hosts we also recorded this for each parameter set. The PRCC could then be calculated for each parameter based on the 100 parameter sets; this does not greatly influence the computational time. The PRCC for the total number infected was calculated as 0.98, -0.004 and 0.86 for β, σ and μ, respectively. Here, increasing β and/or μ significantly increases the total number infected ($p < 1\text{x}10^{-5}$), whereas σ does not ($p = 0.96$). None of the parameters were found to significantly influence the duration of the outbreak.

APPENDIX D

Other epidemiological model results

D.1 Growth rate model for African horse sickness virus (Morocco, 1989)

D.1.1 Introduction

Growth rate models are based on the deterministic SIR model and aim to predict the basic reproduction number, R_0, from the early dynamics of an outbreak. They assume that the time spent in the infectious stage follows an exponential distribution with mean infectious period:

$$D = 1/\mu, \tag{D.1}$$

where μ is the recovery rate for the disease. It is also assumed that the population is closed, therefore the population size (N) will not change during the initial stages of the outbreak ($N'(t) = 0$).

These models attempt to approximate the growth rate r. R_0 can be derived from the growth rate. In the case where the latent period is short compared to the infectious period and the infectious period is assumed to follow an exponential distribution:

$$R_0 = 1 + rD, \tag{D.2}$$

where D is the average infectious period. However if the latent period is not sufficiently short compared to the infectious period and we assume both the latent and infectious periods follow an exponential distribution

$$R_0 = (1 + rD)(1 + rD'), \tag{D.3}$$

where D' and D are the average length of the latent and infectious periods, respectively. If the latent and infectious periods are unknown but assumed to follow an exponential distribution and the serial interval is known

$$R_0 = 1 + rT_s, \tag{D.4}$$

where T_s is the serial interval. Given the case where $R_0 = 1 + rD$ it can be shown that

$$R_0 = 1 + \frac{log(2)}{T_d}D, \tag{D.5}$$

where T_d is the doubling time.

Using the growth rate model approach we attempted to estimate the growth rate for African horse sickness virus (AHSV) during the initial stages of this outbreak. From this we estimated the basic reproduction number R_0.

D.1.2 Methods

Multiple methods have been suggested for calculating the growth rate. We will consider the exponential growth model, generalised growth model, Richards model and generalised Richards model.

Exponential growth model The simplest form of growth model is the exponential growth model. During the early stages of an outbreak it is assumed that

$$I(t) \approx I(0)e^{rt}, \tag{D.6}$$

where $I(t)$ is the number of infectious individuals at time t and r is the growth rate of the epidemic. Taking the logarithm of both sides of Equation D.6 yields:

$$log(I(t)) = log(I(0)) + rt. \tag{D.7}$$

Therefore, we can model this using a linear regression commonly denoted as $y = mx + c$, where $y = ln(I(t))$, $x = t$, $m = r$ and $c = ln(I(0))$.

Generalised growth model There are some issues with the exponential growth model, for example it does not consider under-reporting or reporting delays and biases. We may also not observe exponential growth. Factors such as spatial heterogeneity and behavioural changes mean that epidemic growth could be sub-exponential (slower than exponential). This can be described by the generalised growth model [Viboud et al., 2016]. This model is given by:

$$\frac{dC(t)}{dt} = C'(t) = rC(t)^p, \tag{D.8}$$

where $C(t)$ is the cumulative number of cases, $rC(t)^p$ is the incidence curve, r is the growth rate and $p \in [0, 1]$ is the growth scaling parameter. This growth scaling parameter allows us to determine whether the growth is exponential ($p = 1$), sub-exponential ($0 < p < 1$) or whether there is constant incidence over time ($p = 0$) [Chowell, 2017].

Richard's model Richard's model has been fitted to a range of logistic-type epidemic curves; including outbreaks of vector-borne diseases [Richards, 1959, Dinh et al., 2016]. The model for the number of new infected cases at time t is given by the differential equation

$$\frac{dC(t)}{dt} = rC(t)\left(1 - \left(\frac{C(t)}{K}\right)^a\right), \tag{D.9}$$

where r is the growth rate, K is the size of the epidemic and a measures the extent of derivation from the S-shaped dynamics commonly observed in epidemiological models. During the early stages of an outbreak when $C(t)$ is significantly smaller than K, the Richard's model assumes the growth to be exponential. This growth then dampens over time.

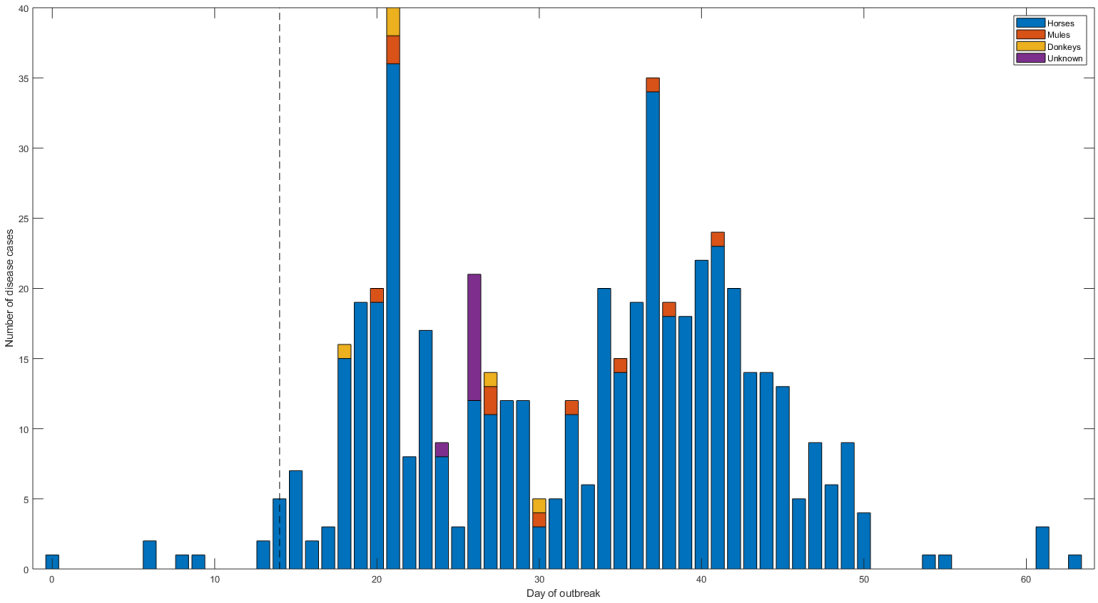

Figure D.1: Time-series of reported AHSV cases in Morocco in 1989 by species. Data before the black dashed line was used to estimate the growth rate.

Generalised Richard's model We can adapt the Richard's model, similarly to the exponential growth model, to account for sub-exponential growth dynamics by incorporating a growth scaling parameter p. This gives a model for the number of new infected cases at time t given as

$$\frac{dC(t)}{dt} = rC(t)^p\left(1 - \left(\frac{C(t)}{K}\right)^a\right), \tag{D.10}$$

where r, K and a are the same parameters as those used in the Richard's model. Similarly to the generalised growth model the growth scaling parameter allows us to determine whether the growth is exponential ($p = 1$), sub-exponential ($0 < p < 1$) or whether there is constant incidence over time ($p = 0$) [Chowell, 2017].

The parameter identifiability of the models was checked using the parametric bootstrap method described in Appendix B. Here $S = 200$ parameter estimates were performed were all parameters in the model where jointly estimated. The initial guesses for parameters r, a and K where sampled using Latin hypercube sampling (LHS) from a uniform distribution with minimum 0 and maximum of double the true estimated value (Table D.1). The initial guesses for the parameter p were also sampled using LHS from a uniform distribution with minimum 0, however the maximum was set to 1 as this is the maximum value that p can take.

Sensitivity analysis was performed using the heat map LHS-PRCC method described in Appendix C. Here 100 parameters are sampled using LHS from a uniform distribution with minimum and maximum values equal to the minimum and maximum estimates from the $S = 200$ parameter estimates generated during the parameter identifiability analysis. This analysis was completed 100 times and the mean is shown.

D.1.3 Results

Figure D.1 shows the incidence per day. We observe an initial increase in cases, followed by a decrease in incidence and then another increase. We consider the early stages of the outbreak the first 15 days; shown by the dashed line on the graph.

The exponential growth model (EGM), generalised growth model (GGM), Richards model (RM) and generalised Richards model (GRM) were fitted to the data; the estimated parameters for each model are shown in Table D.1. We observe that all models estimated the growth rate, r to be 0.1935. Whenever the model included a growth scaling parameter, p, it was estimated to be 1. From this we can therefore conclude that the growth during the initial stages of the outbreak was not sub-exponential. The estimates for a and K were different for the Richard's and generalised Richards models. However, as p was estimated to be 1 these models are the same, therefore we would expect all of the other parameters to be estimated the same.

The mean, median and 5–95% confidence intervals of parameter estimates derived from the parameter identifiability analysis are shown in Table D.1. For all four models the median estimate from the parameter identifiability analysis was closer to the original estimate from the data than mean estimate. We also observe larger confidence intervals for models with more parameters. The large confidence intervals for the RM and GRM offer an explanation to why the estimates of a and K were different for these two models: unidentifiable parameters.

Model	Estimate	r	p	a	K
EGM	True	0.1935	-	-	-
	Mean	0.1910	-	-	-
	Median	0.1940	-	-	-
	5- 95% CI	(0.16,0.22)	-	-	-
GGM	True	0.1935	1	-	-
	Mean	0.2488	0.8745	-	-
	Median	0.2094	1	-	-
	5- 95% CI	(0.17,0.40)	(0.45,1)	-	-
RM	True	0.1935	-	10.04	99.65
	Mean	0.3872	-	17.82	113.39
	Median	0.2036	-	13.11	103.09
	5- 95% CI	(0.16,1.28)	-	(0.04,56)	(14,273)
GRM	True	0.1935	1	16.89	748.24
	Mean	0.2458	0.8754	19.36	789.15
	Median	0.2105	1	18.76	772.54
	5- 95% CI	(0.18,0.42)	(0.48,1)	(6.13,32.82)	(20,1479)

Table D.1: The parameter estimates and confidence intervals (CIs) from the parameter identifiability analysis for the exponential growth model (EGM), generalised growth model (GGM), Richard's model (RM) and generalised Richard's model (GRM) for African sickness virus during the early stages of the 1989–1991 outbreak in Morocco. The True value represents the original estimate to the data.

Results from the sensitivity analysis are shown in Figure D.2. Neither a or K were found to significantly influence the number of incidences at any of the time-points used to fit the models. Both r and p were strongly correlated with incidences for every time-point after $t = 2$. The differential equations for the RM and GRM (Equations D.9 and D.10, respectively) both contain the multiplying term $1 - (C(t)/K)^a$. Therefore, as K gets large the value of this term will approach one and the value of a will have less impact on the final result. Therefore, this could cause structural identifiability issues for large values of K. As there are identifiability issues for a and K in the RM and GRM, we discard these models for this data set and conclude that the EGM is most suitable.

Using the EGM, the growth curve was fitted to the initial stages of the outbreak (Figure D.3a). The residuals between the best fit curve and the data are shown in Figure D.3b. Initially there are five positive residuals; these are caused by the gap between the first and second cases. After the second case on day 6 these residuals are distributed randomly. Results from the parameter identifiability analysis are shown in Figures D.3 c and d. We observe again here that the median estimate (green line) is closer to the true estimate (red line) than the mean estimate (black line). All the data points are within

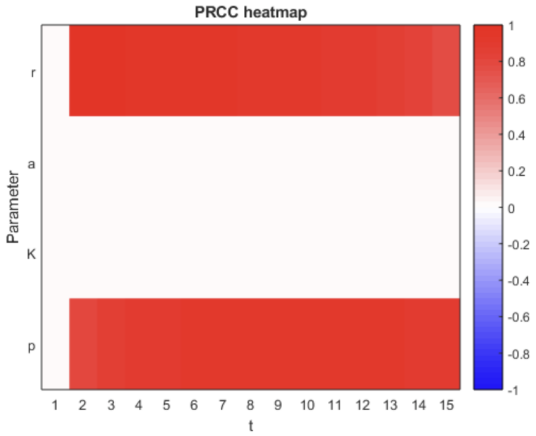

Figure D.2: Results from the LHS-PRCC heatmap sensitivity analysis method applied to the generalised growth model for the early stages (first 15 days) of the African horse sickness virus Morocco 1989–1991 outbreak. A Student's t-test was performed and a and K were not found to significantly influence the number of new cases at any time-point (p>0.05).

the 0.025 and 0.975 quantiles of the estimated epidemic curves. These curves are used along with the their corresponding parameter estimates to forecast the number of daily incidences for the following 7 days (Figure D.3e). We observe that most of the data points follow these trajectories, however many are near the edges of our predictions, perhaps indicating an underestimation of the potential for this disease to spread.

Using our calculated growth rate, $r = 0.1973(0.16 - 0.22CI)$ we can estimate R_0 given the mean infectious period, D. Previously (Chapter 3), this has been estimated as 3.9 (95% CI: 2.1-6.9) days for hosts that die and 8.7 (95% CI: 5-11) days for hosts that survive. As the latent period was estimated as 4.6 (95% CI: 2-8.3) days this is not sufficiently short compared to the infectious period (especially for dying hosts) so this is included in our estimation of $R- = 0$. Therefore, R_0 is calculated using Equation D.3 yielding 3.38 and 5.18 for hosts that die or survive, respectively.

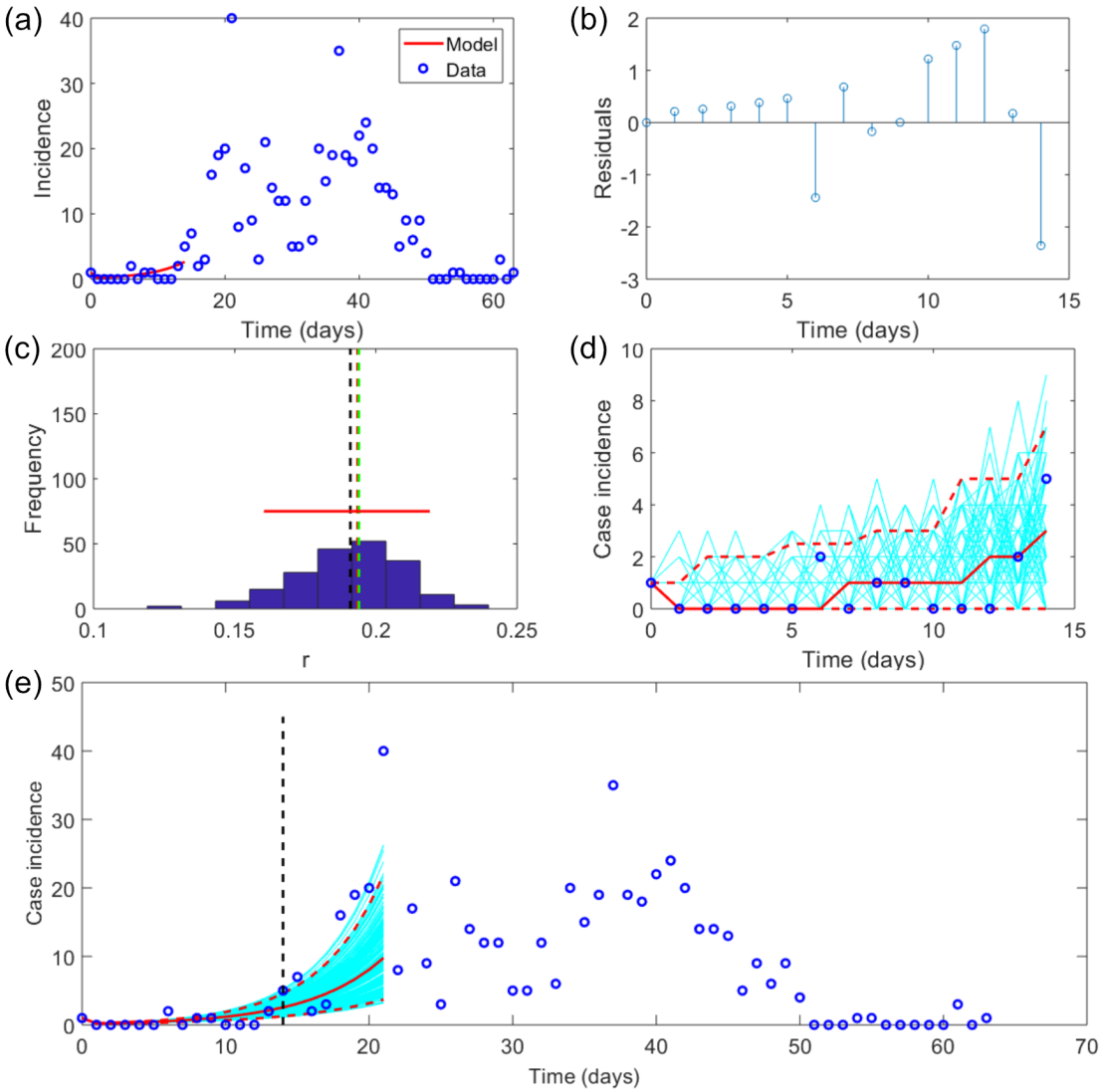

Figure D.3: Results from fitting the exponential growth model (EGM) for the early stages of the 1989-1991 African horse sickness outbreak in Morocco. (a) The model fitted to the data for the first 14 days of the outbreak. (b) Residuals from the curve to the data. (c) Histogram of the $S = 200$ estimated values of r from the parameter identifiability analysis. The horizontal red line represents the 5-95% confidence intervals. The dashed vertical red, black and green lines represent the true, mean and median estimates, respectively. (d) The $S = 200$ curves generated during the parameter identifiability analysis. the blue dots represent the data. The full red line represents the median of the $S = 200$ curves and the red dashed lines represent the 0.975 and 0.025 quantiles. (e) Forecast 7 days beyond the data for the $S = 200$ curves generated during the parameter identifiability analysis. The black dashed line represents where the forecasting begins and all other lines are as in (d).

D.2 Comparing the vector capacity and basic reproduction number of African horse sickness, bluetongue and Schmallenberg viruses in the UK

D.2.1 Introduction

A mechanistic model examines the individual parts of a complex system in order to better understand it. Parameters represent biological processes which can be coupled together to make predictions. Typically, these are used in epidemiology to forecast the risk of disease spread by calculating R_0. They can therefore be used to create risk maps considering the climate and landscape. Brand and Keeling [2017] suggested a novel model for the impacts of temperature variation on transmission dynamics of BTV. We applied the method described by Brand and Keeling [2017] to compare the replication rates, C and R_0 for bluetongue virus (BTV), Schmallenberg virus (SBV) and African horse sickness virus (AHSV). Here R_0 and the vector competence, defined as the contribution to R_0 from one day of biting, are estimated.

D.2.2 Methods

The mean daily temperatures were taken from the Hadley Centre Central England Temperature (HadCET) dataset [Parker et al., 1992], which are representative of a triangular region of the UK enclosed by Lancashire, London and Bristol. The data-sets for the years 2014–2018 were extracted. The temperature-dependent bite rate ($a(T)$), vector mortality rate ($\lambda_V(T)$) and extrinsic incubation period (EIP) completion rate ($\upsilon(T)$) are calculated for each day. Parameters for the bite rate and vector mortality rate are the same for all viruses, as they are spread by the same insects. The same seasonal variation ($\theta(t)$) in abundance of vectors is also used (Equation 1.9).

The Carpenter et al. [2011] parameters and model (Equation 1.3) for the EIP for BTV and AHSV were used, as well as the probability of transmission from host to vector (p_H). The AHSV parameters where $\alpha = 0.017$, $T_{min} = 12.7$, $p_H = 0.52$ and $k = 14.4$. Carpenter et al. [2011] calculated parameters from data from five BTV studies. Our parameters were calculated as the average of the three *C. sonorensis* experiments giving $\alpha = 0.019$, $T_{min} = 12.63$ and $p_H = 0.13$, as this was the species used to obtain the result for AHSV. For BTV the shape parameter (k) for the EIP varied across a large range (7.7–116.2) for the different strains. BTV-9 had a similar shape parameter to AHSV-4, which was the only strain of AHSV tested. Gubbins et al. [2014] attempted to calculate the EIP for Schmallenberg virus (SBV) using outbreak data and approximate Bayesian computation. It was predicted that $\alpha = 0.03$, $T_{min} = 12.35$ and $p_H = 0.14$. There is currently no estimated value available for k for SBV. Therefore, it was decided to only include the Markovian and deterministic solutions. The intermediate model is likely to give an intermediate result, as shown in Brand and Keeling [2017] and is dependent on the uncertain parameter k, whereas the Markovian and deterministic models are not dependent on k and are likely to give an upper and lower bound, respectively. The parameters 'probability of transmission from vector to host' (p_V) and 'recovery rate' ($1/\mu$) for each virus and host species are given in Table D.2.

Virus	Host species	p_V	$1/\mu$	Ref.
BTV	cattle	0.9	20.6	[Jones et al., 2019, Gubbins et al., 2007]
BTV	sheep	0.9	16.4	[Jones et al., 2019, Gubbins et al., 2007]
AHSV	horse	0.77	4.4	[Backer and Nodelijk, 2011]
SBV	cattle	0.76	3.04	[Gubbins et al., 2014]
SBV	sheep	0.76	4.37	[Gubbins et al., 2014]

Table D.2: The parameters for the transmission probability from vector to host and recovery rate for bluetongue virus (BTV), African horse sickness virus (AHSV) and Schmallenberg virus (SBV) and their host species.

D.2.3 Results

The temperature data and the corresponding EIP completion rate for each virus is shown in Figure D.4. We observe that when temperatures are higher, the EIP for all viruses is shorter. During some colder days the EIP completion rate is 0 for all viruses. SBV appears to have a larger EIP completion rate than AHSV and BTV most days, with AHSV and BTV having similar EIP completion rates.

We observe (Figure D.4) that the Markovian model produces the largest estimations for the vector competence and R_0 for all viruses. AHSV had a larger vector competence than BTV for all models, this is due to the larger probability of host to vector transmission. AHSV and BTV have similar estimations of R_0 for both models, with estimations for BTV consistently being slightly larger. This is because BTV has a longer infectious period. Despite the vector competence for SBV being between AHSV and BTV for the deterministic model and similar to BTV for the Markovian model, the estimates for R_0 are lower than AHSV and BTV for all models. This is because SBV parameters included a shorter infectious period than BTV and less probability of transmission from host to vector than SBV. R_0 exceeds 1 for AHSV and BTV in the deterministic and Markovian models and SBV in the Markovian model.

D.2.4 Discussion

Results suggest that AHSV could have the same R_0 as BTV in the UK, however this is likely to be dependent on host availability. Previous outbreaks of BTV in the UK indicate that AHSV would therefore be a threat in this region. Values of R_0 for SBV were rarely above 1. These results could therefore be an underestimation. However, as SBV was parameterised differently to AHSV and BTV (without experimental data) this could be causing this difference.

Updating the estimations for the infectious periods of hosts with AHSV to values identified in Chapter 3 could improve estimates.

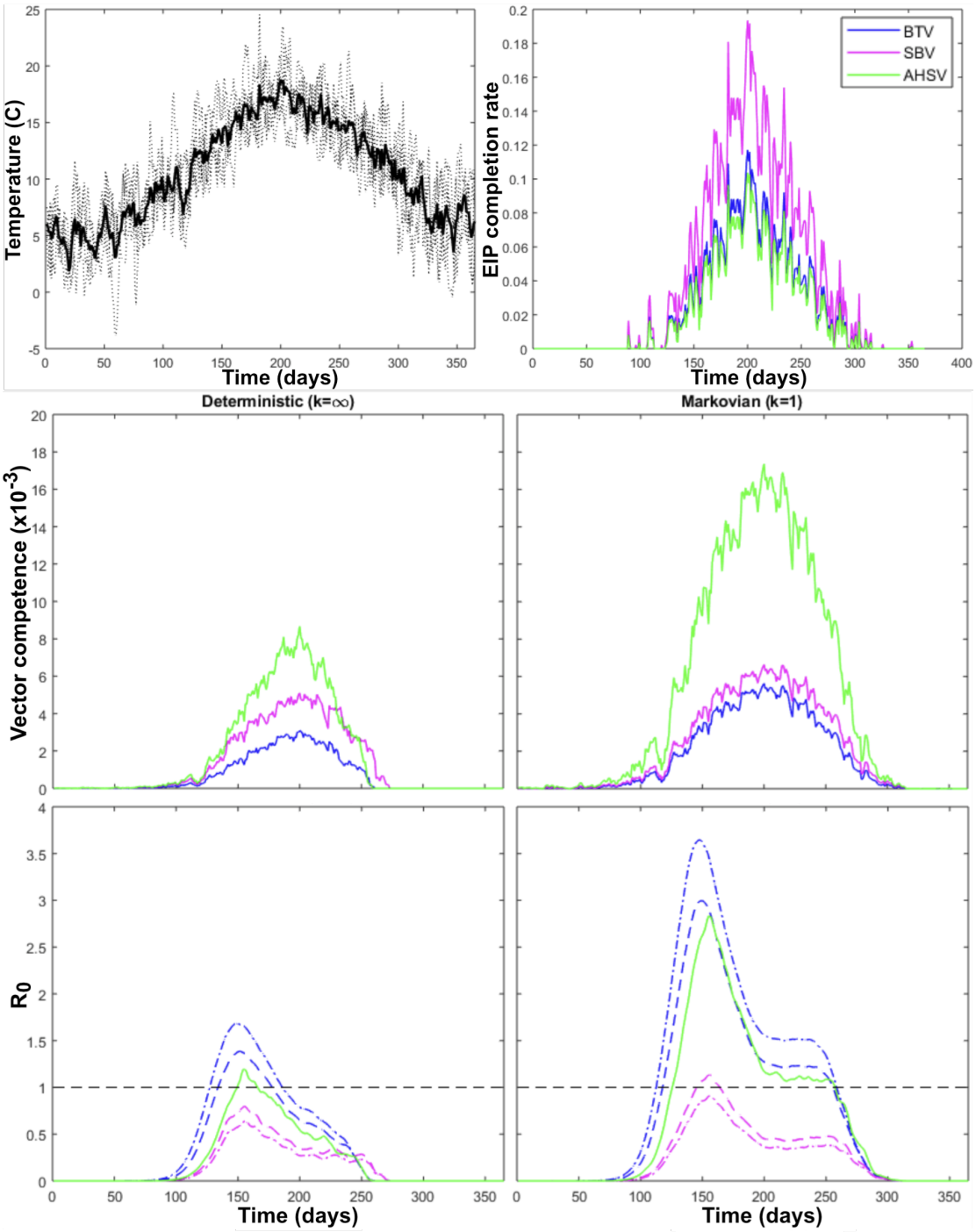

Figure D.4: (Top left) The average daily temperature extracted from the Hadley Centre Central England Temperature (HadCET) dataset for the years 2014–2018 (dotted lines) and their mean (solid line). (Top right) The mean daily EIP completion rate for bluetongue virus (BTV), Schmallenberg virus (SBV) and African horse sickness virus (AHSV). Comparison of the vector competence (middle) and basic reproduction number (bottom) for BTV, SBV and AHSV in the UK using the deterministic (left) and Markovian (right) methods of the mechanistic model described by Brand and Keeling [2017]. Coloured lines represent the mean of the estimations between 2014 and 2018 and the dashed black line represents $R_0 = 1$.

APPENDIX E

Re-parameterisation of a mathematical model of African horse sickness virus using data from a systematic literature search: Supplementary material

Re-parameterisation of a mathematical model of African horse sickness virus using data from a systematic literature search

Emma L. Fairbanks[1], Marnie L. Brennan[1], Peter P.C. Mertens[1], Michael J. Tildesley[2], and Janet M. Daly[1]

[1]School of Veterinary Medicine and Science, University of Nottingham, Loughborough. LE12 5RD, UK
[2]The Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research, School of Life Sciences and Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

Supplementary file 1: Updated model parameters

Table S1 details the vector parameters previously used by Backer & Nodelijk (2011) and the modified values from the literature. Unchanged parameters are also given in this table.

The temperature-dependent rates are calculated for 18°C; to represent the average temperature of equine dense areas in the UK in August. Such areas with large horse populations and a seasonal increased abundance of midges at this time are when an African horse sickness virus outbreak is likely to cause the most impact. When considering an outbreak in a different climate these parameters may need to be adjusted.

There are limited data available for the temperature dependent parameters affecting the transmission of AHSV. For example, an initial search of PubMed using the terms 'African horse sickness' AND 'extrinsic incubation periods' yielded only five results. Of these, two were studies on different viruses (bluetongue virus and equine encephalosis virus (Van Der Saag et al., 2017; Venter et al., 1999)). Carpenter et al. (2011), published shortly after Backer and Nodelijk published their model in 2011, reanalysed data from Wittmann et al. (2002) (which was also found in the search) using novel statistical methodology. The fifth search result (Sánchez-Matamoros et al., 2016) was a mathematical modelling study that referenced Carpenter et al. (2011). Carpenter et al. (2011) derived the Extrinsic incubation period (EIP) and probability of transmission from host to vector from AHSV-4 infection experiments in *Culicoides sonorensis*.

In an attempt to find additional articles quantifying the probability of transmission from host to vector, the search terms 'African horse sickness' AND ('vector competence' OR susceptibility') were used to search PubMed. After screening the titles and abstracts, apart from the original Wittmann et al. (2002) article and the Carpenter et al. (2011) article reanalysing the data, four other articles were found (Venter et al., 2000, 2009, 2010; Venter & Paweska, 2007). These articles all compare the susceptibility of African *Culicoides* species to AHSV. Findings from these studies are highly variable between species (Venter et al., 2000, 2009; Venter & Paweska, 2007), populations (Venter et al., 2009), serotypes (Venter et al., 2009, 2010), isolates (Venter et al., 2009, 2010; Venter & Paweska, 2007) and seasons (Venter et al., 2009). The species of *Culicoides* with the estimated largest probability of being infectious within 10 days at 23.5°C of a viraemic blood meal and this probability in these studies by serotype are given in Table S2. In this study, we use the parameter suggested by Carpenter et al. (2011) (0.52) as it was parameterised with the EIP used in the model and for the same species as other parameters used in the model. The model is also simulated for a temperature of 18°C, which is within the span of temperatures examined in the study (15–30°C).

This trend of limited data availability for the dynamics of AHSV and *Culicoides* continued in the search for other parameters. Other *Culicoides*-borne disease models parameterised the blood feeding interval, assumed to be the time between blood feeding and oviposition, function derived by Mullens et al. (2004) for C. *sonorensis* (Gubbins et al., 2008, 2014; Haider et al., 2019). The updated vector lifespan used was derived by Gerry & Mullens (2000) who attempted to quantify how the expected lifespan of a *C. sonorensis* varies according to tem-

Table S1: Updated model parameters from other literature. The temperature-dependent rates are calculated for 18°C

Parameter	Symbol	Previous value		Updated value*	
		Default value	5-95% range	Default value	5-95% range
Blood feeding interval	$1/a$	7.5	4.7-17.7	6.3[1]	_
EIP	$1/\upsilon$	16	9.2-48	10.9[2]	_
No. stages	k	10		_	
Vector life-span	$1/\mu_V$	22	16-31	11.9[3]	_
Transmission probability host to vector	p_H	0.04	0.01-0.1	0.52[2]	0.45 – 0.59[2]
Transmission probability vector to host	p_V	0.77	0.5-0.95	_	_
Vector:host ratio	ρ	226	1-4219	_	_
Initial number of hosts	N_H	66	32-100	_	_

. Time periods are given in days. The parameter values used in Backer & Nodelijk (2011) are described in the previous value column and the parameter values used in our model are described by the updated value column. Where the updated column is blank no new information on these parameters was found. * [1] = updated using Mullens et al. (2004), [2] = updated using Carpenter et al. (2011), [3] = updated using Gerry & Mullens (2000).

Table S2: Species of Culicoides with the largest probability of being infected within 10 days at 23.5 °C of a viraemic blood meal.

Serotype	*Culicoides* species	Probability	Reference
AHSV-1	C. *leucostictus*	0.08	Venter et al. (2010)
AHSV-2	C. *leucostictus*	0.38	Venter et al. (2010)
AHSV-3	C. *bolitinos*	0.17	Venter et al. (2009)
AHSV-4	C. *imicola*	0.05	Venter et al. (2010)
AHSV-5	C. *bolitinos*	0.21	Venter et al. (2000)
AHSV-6	C. *imicola*	0.41	Venter et al. (2009)
AHSV-7	C. *imicola*	0.33	Venter et al. (2009)
AHSV-8	C. *imicola*	0.27	Venter et al. (2000)
AHSV-9	C. *zulluensis*	0.20	Venter et al. (2009)

perature using data from a dairy farm in California, USA. The Backer and Nodelijk paper (2011) mentioned this study but suggested that the mortality for midges in the field was significantly higher than that seen in lab experiments by Wittmann et al. (2002), the value they used in their model. This leads to reduced transmission when the Gerry & Mullens (2000) parameters are used However, the midge mortality rate may be higher in the field due to environmental effects. The PubMed search 'Culicoides' AND ('survival' OR 'mortality') did not identify any other studies of interest.

Supplementary file 2: Mathematical model description

Let $S_H(t)$, $E_H(t)$, $I_H(t)$, $R_H(t)$ and $D_H(t)$ denote the number of susceptible, latent, infectious, recovered and dead horses at time t. We also denote the total number of alive horses as N_H. All latently infected horses transition to infectious stage I_H^1 when they become infectious. From here they can either transition to D_H (dying hosts) or transition to infectious stage I_H^2 followed by R_H (recovering hosts). The recovered class is not considered in vector populations; this is because of their short lifespans; therefore we denote $S_V(t)$, $E_V(t)$ and $I_V(t)$ as the number of susceptible, latent and infectious vectors. In the model, vector mortality and birth also occur at an equal rate, μ_V, which is dependent on temperature; this allows us to consider the total number of vectors as a constant, denoted N_V, determined by the vector:host ratio. We will assume that all vectors are born susceptible.

The rate at which new infections occur depends on the bite rate of midges, the transmission probabilities and the number of individuals in the S_H, I_H, S_V and I_V classes. The bite rate $(1/\epsilon)$ depends on temperature. The transmission probabilities from host-to-vector and vector-to-host, p_H and p_V, respectively, is the probability that an infectious bite results in a susceptible individual becoming infected. The infection rates for susceptible hosts and vectors are therefore given by

$$p_H \epsilon \frac{N_V}{N_H} \frac{I_V}{N_V} = p_H \epsilon \frac{I_V}{N_H}, \tag{1}$$

and

$$p_V \epsilon \frac{I_H}{N_H}, \tag{2}$$

where N_V/N_H is the vector to host ratio (ρ), I_H/N_H is the proportion of hosts infected and I_V/N_V is the proportion of vectors infected.

The rate latent midges become infected is dependent on temperature and given by the extrinsic incubation period (EIP): $1/\upsilon$. The simplest way to introduce recovery is to assume the infected host recovers at a constant rate γ, which is the inverse of their infectious period.

The adapted deterministic ODE system is given by:

$$\frac{dS_H}{dt} = a p_V \frac{I_v S_H}{N_H}, \tag{3}$$

$$\frac{dE_{H,1}}{dt} = a p_V \frac{I_V S_H}{N_H} - l\epsilon E_{H,1}, \tag{4}$$

$$\frac{dE_{H,i}}{dt} = l\epsilon E_{H,i-1} - l\epsilon E_{H,i}, \text{ for } 2 \leq i \leq l, \tag{5}$$

$$\frac{dI^1_{H,1}}{dt} = l\epsilon E_{H,l} - n^1\gamma^1 I^1_{H,1}, \tag{6}$$

$$\frac{dI^1_{H,i}}{dt} = n^1\gamma^1 I^1_{H,i-1} - n^1\gamma^1 I_{H,i}, \text{ for } 2 \leq i \leq n^1, \tag{7}$$

$$\frac{dI^2_{H,1}}{dt} = (1-m_H) n^1\gamma^1 I^1_{H,n^1} - n^2\gamma^2 I^1_{H,1}, \tag{8}$$

$$\frac{dI^2_{H,i}}{dt} = n^2\gamma^2 I^2_{H,i-1} - n^2\gamma^2 I^2_{H,i}, \text{ for } 2 \leq i \leq N^2, \tag{9}$$

$$\frac{dD_H}{dt} = m_H n^1 \gamma^1 I^1_{H,n^1}, \tag{10}$$

$$\frac{dR_H}{dt} = n^2\gamma^2 I^2_{H,n^2}, \tag{11}$$

$$\frac{dS_V}{dt} = \mu_V N_v - ap_H \frac{S_V(I^1_H + I^2_H)}{N_H} - \mu_V S_V, \tag{12}$$

$$\frac{dE_{V,1}}{dt} = ap_H \frac{S_V(I^1_H + I^2_H)}{N_H} - (k\upsilon + \mu_V)E_{v,1}, \tag{13}$$

$$\frac{dE_{V,i}}{dt} = k\upsilon E_{V,i} - (k\upsilon + \mu_V)E_{V,i}, \text{ for } 2 \leq i \leq k, \tag{14}$$

$$\frac{dI_V}{dt} = k\upsilon E_{V,k} - \mu_V I_V, \tag{15}$$

where the nature of the parameters are described in Tables 5 and S1. The overall recovery rate from the first and second infectious classes are calculated as

$$\gamma^1 = \frac{1}{T^1_{inf}} \text{ and } \gamma^2 = \frac{1}{T^2_{inf} - T^1_{inf}}, \tag{16}$$

respectively.

The latent period of hosts, infectious period of hosts, and EIP of vectors are divided into multiple stages; this allows them to have a gamma distribution (Lloyd, 2001). Here, for example the latent stage (E_H) is subdivided into i stages; each with mean transition rate ϵ/i, where ϵ is the transition rate from stage E_H to stage I^1_H. Using this method, $E_H(t)$ is replaced by a series of i stages $E^1_{H,1}(t)$, $E^1_{H,2}(t)$, $\ldots$,$\mathrm{E}^1_{H,i}(t)$. When individuals first transition from the previous stage, the individuals will enter the first stage $E^1_{H,1}(t)$. Then, once each stage is completed, they will continue to pass through all stages successively until they complete the $E^1_{H,i}(t)$ stage.

Supplementary file 3: Sensitivity analysis

The Latin hypercube sampling partial rank correlation coefficient method (LHS-PRCC method) was used for the sensitivity analysis. This method was described in Blower & Dowlatabadi (1994). If the PRCC is close to 0, the parameter is weakly (or not at all) correlated with the output parameter. As the PRCC approaches 1 or -1 it represents a positive or negative correlation, respectively, with 1 and -1 reflecting a perfect correlation. This method can also be applied to each time-step in order to see the sensitivity of the parameters over the time-course; here we refer to this as the heat map method.

We perform the calculation of the PRCC 100 times for 100 sets of parameters each time. The median PRCC and its variation are examined using a box-plot. The duration of the outbreak is considered to be when it is $< 10^{-3}$ from its final steady state. This is due to the numerical methods resulting in minute changes in the populations as the steady state is approached. For the heat map method, the PRCC is calculated only once per time-step and is performed 100 times for 100 parameters sets; the value plotted is the mean PRCC of the outputs. The heat maps will vary from white to red for 0 to 1 and from white to blue from 0 to -1. Therefore the more positively correlated a parameter is the more red it will be, and the more negatively correlated a parameter is the more blue it will be. If a result was not found to be significant through the Student's t-test ($p > 0.05$) it was coloured white. Matlab was used for calculations and to produce figures, LHS was performed using the *lhsdesign* function. Here the ODE solver *ode23s* was used, alternatively to *ode45*, to resolve stiffness issues for some sampled parameter sets.

Parameters were all selected for a uniform distribution (upper and lower bounds shown in Table S3). The upper and lower bounds for the latent and infectious periods of hosts were set to the maximum and minimum values found in the systematic review. However, this caused numerical issues when the length of the infectious period selected for dying hosts was longer than that of surviving hosts. Therefore, the infectious period of surviving hosts was set to the infectious period of the dying hosts multiplied by 1.2, as the mean infectious period of surviving hosts found in the systematic review was 1.2 times longer than that of dying hosts. The transmission probability from host to vector was set to the 5–95% range found in Carpenter et al. (2011). As the parameters for the transmission probability from vector to host, vector to host ratio and the initial number of hosts were not updated, the 5–95% range from Backer and Nodelijk (2011) were used. In Backer and Nodelijk (2011) the host case fatality 5–95% range varied -/+0.27 from the default value. This variance was also applied to the updated case fatality with a maximum of 1. No confidence intervals were given for the blood feeding interval and vector life-span therefore the upper and lower bounds were set to +/-25% of the updated default values. The extrinsic incubation period is calculated using two parameters estimated in Carpenter *at al.* (2011) where the 5–95% confidence intervals are given. In order to find the lower and upper bound for the sensitivity analysis the extrinsic incubation period was calculated for 10000 random values between the 5–95% confidence intervals for these parameters. The upper and lower bounds are then taken from 5% and 95% of these 10000 values, respectively.

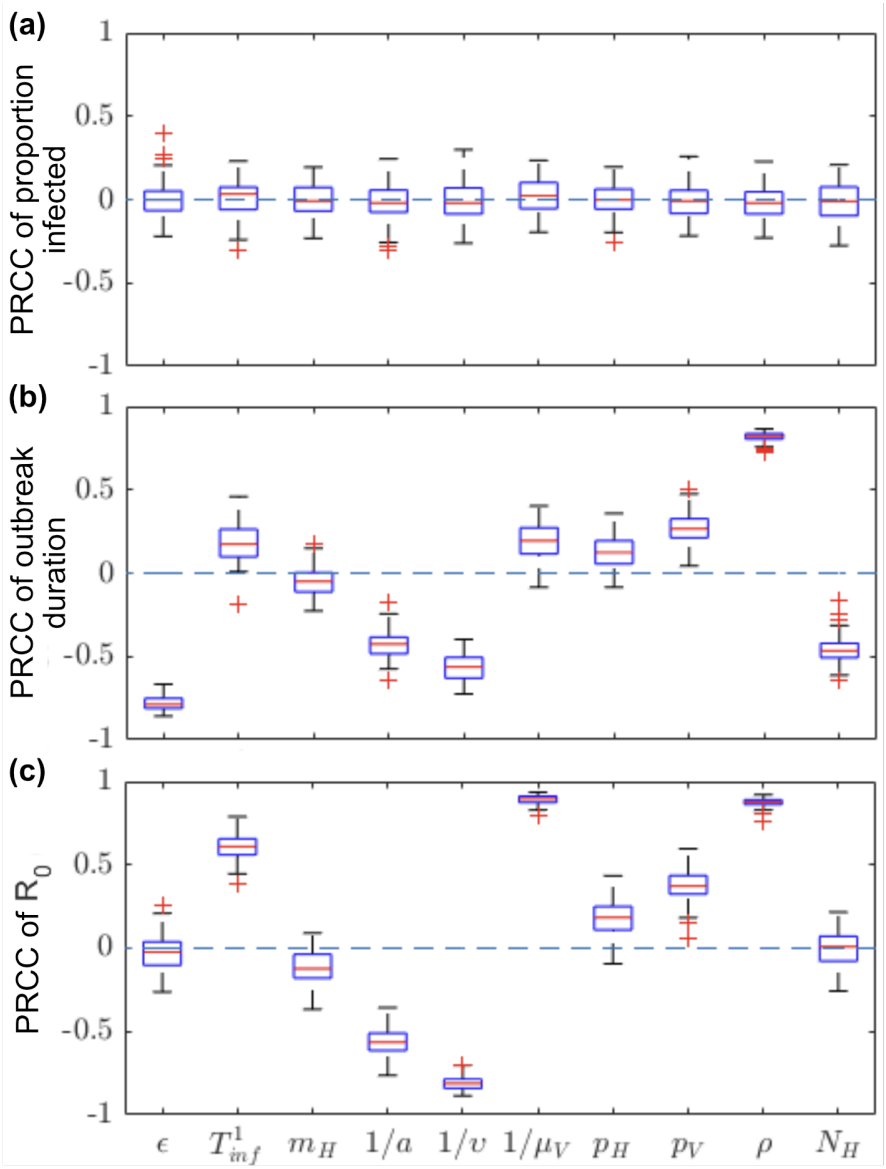

Figure S1: Box and whisker plots of PRCC between the model parameters and the (a) proportion of hosts infected, (b) outbreak duration and (c) R_0. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25^{th} and 75^{th} percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the + symbol.

Table S3: The upper and lower bounds of parameters used for the sensitivity analysis. Time periods are given in days. To avoid numerical errors $T^2_{inf} = 1.2T^1_{inf}$.

Parameter	Lower bound	Upper bound
$1/\epsilon$	2	11
T^1_{inf}	2	7
m_H	0.57	1
$1/a$	4.7	7.9
$1/\upsilon$	8.6	13.5
$1/\mu_V$	8.9	14.9
p_H	0.45	0.59
p_V	0.5	0.95
ρ	1	4219
N_H	32	100

The PRCC method is applied to the duration of the outbreak, total equine infections over the duration of the outbreak and R_0 (Figure S1). Here R_0 is calculated using the equation suggested by Backer & Nodelijk (2011); given as

$$R_0 = \sqrt{a^2 p_H p_V \rho\big(m_H T^1_{inf} + (1 - m_H)T^2_{inf}\big)\left(\frac{k\upsilon}{k\upsilon + \mu_V}\right)^k \frac{1}{\mu_V}}. \tag{17}$$

The heat map PRCC method on the case incidences in each time-step are shown in Figure S2. Table S4 also shows the median PRCC values and significance for the PRCC on the duration of the outbreak, total number of equines infected and R_0. Here two-tailed Student's t-tests were used to determine the p-values.

We observe from Figure S1 and Table S4 that none of the parameters significantly affect the proportion of equids infected. Despite the uncertainty in many parameters, most simulations in the sensitivity analysis resulted in all the equids on a premises becoming infected. The duration of the latent period ($1/\epsilon$) and number of horses on the premises (N_H) are not used in the calculation of R_0, therefore they do not significantly influence its value. The host case fatality (m_H) did not significantly influence R_0, suggesting that culling may not be an effective control strategy. The parameter which most significantly influences the duration of the outbreak is the vector:host ratio. This is associated with shorter outbreaks. This can also be observed in the heat map (Figure S2). The heat map shows that longer host latent periods ($1/\epsilon$), duration between vector bites ($1/a$) and extrinsic incubation periods ($1/\upsilon$) increase the duration of the outbreak (negatively associated with case incidence during the early stages). However, the host infectious period (T_{inf}), vector life-span ($1/\mu_V$), host

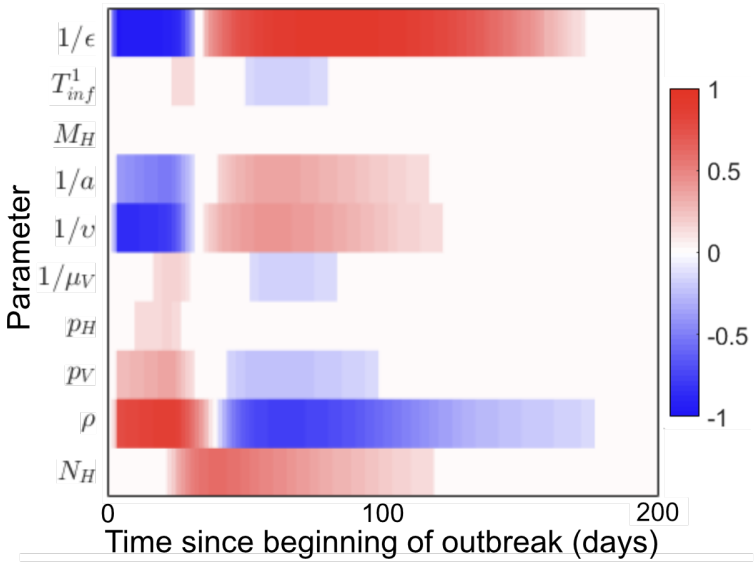

Figure S2: Heatmap of the PRCC of case incidence each timestep.

Table S4: Results from the sensitivity analysis ; the PRCC values and t-values of parameters on the proportion of equines infected, duration of the outbreak and R_0. The PRCC values are given to 3 decimal places. The t-values are given to 2 decimal places. P-values where calculated using a two-tailed students t-test with 89 degrees of freedom. * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$.

PRCC of:	Proportion infected		Outbreak duration		R_0	
Parameter	Mean	t value	Mean	t value	Mean	t value
$1/\epsilon$	-0.000	-0.00	-0.778	-5.78***	-0.022	-0.22
T^1_{inf}	0.007	0.07	0.180	1.97	0.609	9.64***
m_H	-0.001	-0.01	-0.053	-0.51	-0.121	-1.13
$1/a$	-0.021	-0.21	-0.429	-3.55***	-0.558	-4.43***
$1/\upsilon$	-0.010	-0.10	-0.565	-4.47***	-0.810	-5.96***
$1/\mu_V$	0.027	0.27	0.188	2.07*	0.892	26.91***
p_H	-0.000	-0.00	0.124	1.31	0.184	2.01*
p_V	-0.010	-0.10	0.270	3.12**	0.377	4.73***
ρ	-0.020	-0.19	0.817	18.88***	0.875	24.53***
N_H	-0.010	-0.10	-0.461	-3.78***	-0.001	-0.01

to vector (p_H) and vector to host (p_V) transmission rates, and the vector:host ratio are associated with outbreaks that spread more rapidly upon emergence (positively associated with case incidence for up to approximately the first 40 days).

Supplementary table S5

Table S5: Qualitative synthesis of the studies found reporting AHSV experimental infection in a naive equid. If reported the time until viraemia, onset of clinical signs and death of the equid is given. Where possible the duration of viraemia and clinical signs is also given. *n/a = survived; †= died; ‡= euthanised; †/‡= died or euthanised. _ = data unavailable.

Reference	Inoculation route	Viraemia detection method	Serotype	Equid sp.	Days to: Viraemia	Clinical signs	Death*	Duration of: Viraemia	Clinical signs
Lelli et al. (2013)	Intravenous	PCR/isolation	AHSV-9	Horse	11/12	10	n/a	4	5
				Horse	7/8	8	n/a	5	3
Lulla et al. (2017)	Intravenous	PCR	AHSV-4	Horse	4	6	10‡	6	4
Alberca et al. (2014)	Intravenous	PCR/isolation	AHSV-9	Horse	3/3	3	5†	2	2
				Horse	3/3	3	6†	3	3
				Horse	3/3	3	6†	3	3
von Teichman et al. (2010)	Intravenous	Isolation	AHSV-5	Horse	3	_	‡		
			AHSV-6	Horse	5	_	‡		
			AHSV-8	Horse	5	_	‡		
			AHSV-9	Horse	5	_	‡		
Guthrie et al. (2009)	Intravenous	PCR/isolation	AHSV-4	Horse	8/8	_	n/a		
Scanlen et al. (2002)	Subcutaneous	Isolation	AHSV-5	Horse	3	4	7†	4	3
				Horse	4	4	8†	4	4
Du Plessis et al. (1998)	Subcutaneous	n/a	AHSV-5	Horse	_	_	7‡		
Roy et al. (1996)	_	Isolation	AHSV-4	Horse	6	6	10†	4	4

Martínez-Torrecuadrada et al. (1996)	Intravenous	Isolation	AHSV-4	Horse	6		†		
Stone-Marschat et al. (1996)	Intravenous	Isolation	AHSV-4	Horse	3	4	†/‡		
				Horse	3	4	†/‡		
				Horse	3	4	†/‡		
				Horse	3	4	†/‡		
J. House et al. (1994)	Intravenous	Isolation	AHSV-4	Horse	3	5	7†	4	2
				Horse	4	6	8†	4	2
Hassanain (1992)	_	n/a	AHSV-9	Horse	_	_	13†		
Mirchamsy & Taslimi (1968)	Intravenous	n/a	AHSV-9	Horse	_	_	17†		
Ozawa & Bahrami (1966)	Intravenous	n/a	AHSV-9	Horse	_	8	9†		
C. House et al. (1990)	Subcutaneous	n/a	AHSV-1	Horse	_	_	†/‡		
			AHSV-2	Horse	_	_	†/‡		
			AHSV-3	Horse	_	_	†/‡		
			AHSV-4	Horse	_	_	†/‡		
El Hasnaoui et al. (1998)	Subcutaneous	n/a	AHSV-4	Donkey	_	_	_		
				Donkey	_	_	_		
				Donkey	_	_	_		
				Donkey	_	_	_		
				Mule	_	_	_		
				Mule	_	_	_		
				Mule	_	_	_		
van Rijn et al. (2018)	Intravenous	PCR	AHSV-5	Horse	2	3	6†	4	3
				Horse	3	4	8†	5	4

Minke et al. (2012)	Intravenous	PCR	AHSV-4	Horse	8	_	n/a		
Martínez-Torrecuadrada et al. (1997)	Intravenous	n/a	AHSV-4	Horse	_	_	9†		
				Horse	_	_	16†		
				Horse	_	_	n/a		
Alexander & Du Toit (1934)	Intravenous	n/a	_	Horse	_	3	6†		3
Dubourget et al. (1992)	Subcutaneous	Isolation	AHSV-4	Horse	2	_	9†	7	
				Horse	6	_	n/a	5	
Mirchamsy & Taslimi (1964b)	Intravenous	n/a	AHSV-9	Horse	_	_	12†		
J. House et al. (1992)	Intravenous	n/a	AHSV-9	Horse	_	_	7†		
				Horse	_	_	7†		
				Horse	_	_	7†		
Hazrati & Ozawa (1965)	Intravenous	n/a	ASHV-9	Horse	_	_	14†		
				Horse	_	_	n/a		
				Donkey	_	_	n/a		
Whitworth (1930)	_	n/a	_	Horse	_	3	5 ‡		2
Quan et al. (2010)	Intravenous	PCR	AHSV-4	Horse	7	_	_		
Mirchamsy & Taslimi (1964a)	Intravenous	n/a	AHSV-9	Horse	_	_	†		
				Horse	_	_	†		
				Horse	_	_	†		
				Horse	_	_	†		
Ozawa et al. (1970)	_	n/a	AHSV-9	Horse	_	_	9†		
Ozawa et al. (1965)	Intravenous	n/a	AHSV-9	Horse	_	_	14†		

Supplementary table S6

Results from the Kruskal-Wallis test to determine differences between the time until viraemia, onset of clinical signs and death of different serotypes of African horse sickness virus and the method used for their inoculation.

Table S6: The chi-squared value, degrees of freedom and p-values are given. † Serotypes with only one data value are not included.

Time until:	χ^2	df	p-value
Serotype			
Viraemia	4.58	4	0.33
Viraemia †	3.68	2	0.16
Clinical signs	2.09	2	0.35
Death	2.68	2	0.26
Inoculation method			
Viraemia	0.55	1	0.46
Clinical signs	0.08	1	0.78
Death	0	1	1

References

Alberca, B., Bachanek-Bankowska, K., Cabana, M., Calvo-Pinilla, E., Viaplana, E., Frost, L., ... Castillo-Olivares, J. (2014). Vaccination of horses with a recombinant modified vaccinia Ankara virus (MVA) expressing African horse sickness (AHS) virus major capsid protein VP2 provides complete clinical protection against challenge. *Vaccine*, *32*(29), 3670–3674. doi: 10.1016/j.vaccine.2014.04.036

Alexander, R. A., & Du Toit, P. J. (1934). The immunization of horses and mules against horse sickness by means of the neurotropic virus of mice and guinea pigs. *Onderstepoort J Vet Res*, *2*, 375-391.

Backer, J. A., & Nodelijk, G. (2011). Transmission and control of African horse sickness in The Netherlands: a model analysis. *PLoS One*, *6*(8), e23066. doi: 10.1371/journal.pone.0023066

Blower, S. M., & Dowlatabadi, H. (1994). Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int Stat Rev*, 229-243. doi: 10.2307/1403510

Carpenter, S., Wilson, A., Barber, J., Veronesi, E., Mellor, P., Venter, G., & Gubbins, S. (2011). Temperature dependence of the extrinsic incubation period of orbiviruses in Culicoides biting midges. *PloS one*, *6*(11), e27987. doi: 10.1371/journal.pone.0027987

Dubourget, P., Preaud, J., Detraz, N., Lacoste, F., Fabry, A., Erasmus, B., & Lombard, M. (1992). Development, production and quality control of an industrial inactivated vaccine against African horse sickness virus serotype 4. *Bluetongue, African horse sickness and related orbiviruses, CRC Press, Boca Raton*, 874-886.

Du Plessis, M., Cloete, M., Aitchison, H., & Van Dijk, A. (1998). Protein aggregation complicates the development of baculovirus-expressed African horse sickness virus serotype 5 VP2 subunit vaccines. *Onderstepoort J Vet Res*, *65*(4), 321-329.

El Hasnaoui, H., El Harrak, M., Zientara, S., Laviada, M., & Hamblin, C. (1998). Serological and virological responses in mules and donkeys following inoculation with African horse sickness virus serotype 4. *Arch Virol Suppl*, *14*, 29-36. doi: 10.1007/978-3-7091-6823-3\underline{}4

Gerry, A. C., & Mullens, B. A. (2000). Seasonal abundance and survivorship of Culicoides sonorensis (Diptera: Ceratopogonidae) at a southern California dairy, with reference to potential bluetongue virus transmission and persistence. *J Med Entomol*, *37*(5), 675-688. doi: 10.1603/0022-2585-37.5.675

Gubbins, S., Carpenter, S., Baylis, M., Wood, J. L., & Mellor, P. S. (2008). Assessing the risk of bluetongue to UK livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *J R Soc Interface*, *5*(20), 363-371. doi: 10.1098/rsif.2007.1110

Gubbins, S., Turner, J., Baylis, M., Van der Stede, Y., van Schaik, G., Abrahantes, J. C., & Wilson, A. J. (2014). Inferences about the transmission of Schmallenberg virus within and between farms. *Prev Vet Med*, *116*(4), 380-390. doi: 10.1016/j.prevetmed.2014.04.011

Guthrie, A. J., Quan, M., Lourens, C. W., Audonnet, J.-C., Minke, J. M., Yao, J., ... MacLachlan, N. J. (2009). Protective immunization of horses with a recombinant canarypox virus vectored vaccine co-expressing genes encoding the outer capsid proteins of African horse sickness virus. *Vaccine*, *27*(33), 4434-4438. doi: 10.1016/j.vaccine.2009.05.044

Haider, N., Kjær, L. J., Skovgård, H., Nielsen, S. A., & Bødker, R. (2019). Quantifying the potential for bluetongue virus transmission in danish cattle farms. *Sci Rep*, *9*(1), 1–12. doi: 10.1038/s41598-019-49866-8

Hassanain, M. M. (1992). Preliminary findings for an inactivated African horse sickness vaccine using binary ethyleneimine. *Rev Elev Med Vet Pays Trop*, *45*(3-4), 231-234.

Hazrati, A., & Ozawa, Y. (1965). Monovalent live-virus horse-sickness vaccine. *Bull Off Int Epizoot*, *64*, 683-695.

House, C., Mikiciuk, P. E., & Beminger, M. L. (1990). Laboratory diagnosis of African horse sickness: comparison of serological techniques and evaluation of storage methods of samples for virus isolation. *J Vet Diagn Invest*, *2*(1), 44-50. doi: 10.1177/104063879000200108

House, J., Lombard, M., Dubourget, P., House, C., & Mebus, C. A. (1994). Further studies on the efficacy of an inactivated African horse sickness serotype 4 vaccine. *Vaccine*, *12*(2), 142-144. doi: 10.1016/0264-410x(94)90052-3

House, J., Lombard, M., House, C., Dubourget, P., & Mebus, C. A. (1992). Efficacy of an inactivated vaccine for African horse sickness serotype 4. *Bluetongue, African horse sickness and related orbiviruses, CRC Press, Boca Raton*, 891-895.

Lelli, R., Molini, U., Ronchi, G. F., Rossi, E., Franchi, P., Ulisse, S., ... others (2013). Inactivated and adjuvanted vaccine for the control of the African horse sickness virus serotype 9 infection: Evaluation of efficacy in horses and guinea-pig model. *Vet Ital*, *49*(1), 89-98.

Lloyd, A. L. (2001). Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol*, *60*(1), 59–71.

Lulla, V., Losada, A., Lecollinet, S., Kerviel, A., Lilin, T., Sailleau, C., ... Roy, P. (2017). Protective efficacy of multivalent replication-abortive vaccine strains in horses against African horse sickness virus challenge. *Vaccine*, *35*(33), 4262-4269. doi: 10.1016/j.vaccine.2017.06.023

Martínez-Torrecuadrada, J. L., Díaz-Laviada, M., Roy, P., Sánchez, C., Vela, C., Sánchez-Vizcaíno, J. M., & Casal, J. I. (1996). Full protection against African horsesickness (AHS)

in horses induced by baculovirus-derived AHS virus serotype 4 VP2, VP5 and VP7. *J Gen Virol*, *77*(6), 1211-1221. doi: 10.1099/0022-1317-77-6-1211

Martínez-Torrecuadrada, J. L., Díaz-Laviada, M., Roy, P., Sánchez, C., Vela, C., Sanchez-Vizcaino, J. M., & Casal, J. I. (1997). Serologic markers in early stages of African horse sickness virus infection. *J Clin Microbiol*, *35*(2), 531-535. doi: 10.1128/jcm.35.2.531-535.1997

Minke, J. M., Audonnet, J.-C., Guthrie, A. J., MacLachlan, N. J., & Yao, J. (2012). *Vaccine against African horse sickness virus.* Google Patents. (US Patent 8,168,200)

Mirchamsy, H., & Taslimi, H. (1964a). Attempts to vaccinate foals with living tissue culture adapted horse sickness virus. *Arch Inst Razi*, *17*(1), 17-27. doi: 10.22092/ari.1965.108547

Mirchamsy, H., & Taslimi, H. (1964b). Immunization against African horse-sickness with tissue culture adapted neurotropic viruses. *Brit vet J*, *120*(10), 481-486. doi: 10.1016/S0007-1935(17)41556-9

Mirchamsy, H., & Taslimi, H. (1968). Inactivated African horse sickness virus cell culture vaccine. *Immunology*, *14*(1), 81-88.

Mullens, B., Gerry, A., Lysyk, T., & Schmidtmann, E. (2004). Environmental effects on vector competence and virogenesis of bluetongue virus in Culicoides: interpreting laboratory data in a field context. *Vet Ital*, *40*(3), 160-166.

Ozawa, Y., & Bahrami, S. (1966). African horse-sickness killed-virus tissue culture vaccine. *Can J Comp Med Vet Sci*, *30*(11), 311-314.

Ozawa, Y., Hazrati, A., & Bahrami, S. (1970). African horse-sickness live and killed virus tissue culture vaccine. *Arch Inst Razi*, *22*, 103-111.

Ozawa, Y., Hazrati, A., & Erol, N. (1965). African horse-sickness live-virus tissue culture vaccine. *Arch Inst Razi*, *18*, 61-84.

Quan, M., Lourens, C. W., MacLachlan, N. J., Gardner, I. A., & Guthrie, A. J. (2010). Development and optimisation of a duplex real-time reverse transcription quantitative PCR assay targeting the VP7 and NS2 genes of African horse sickness virus. *J Virol Methods*, *167*(1), 45-52. doi: 10.1016/j.jviromet.2010.03.009

Roy, P., Bishop, D. H., Howard, S., Aitchison, H., & Erasmus, B. (1996). Recombinant baculovirus-synthesized African horse sickness virus (AHSV) outer-capsid protein VP2 provides protection against virulent AHSV challenge. *J Gen Virol*, *77*(9), 2053-2057. doi: 10.1099/0022-1317-77-9-2053

Sánchez-Matamoros, A., Sánchez-Vizcaíno, J., Rodríguez-Prieto, V., Iglesias, E., & Martínez-López, B. (2016). Identification of suitable areas for African horse sickness virus infections in Spanish equine populations. *Transbound Emerg Dis*, *63*(5), 564–573. doi: 10.1111/tbed.12302

Scanlen, M., Paweska, J., Verschoor, J., & Van Dijk, A. (2002). The protective efficacy of a recombinant VP2-based African horse sickness subunit vaccine candidate is determined by adjuvant. *Vaccine*, *20*(7-8), 1079-1088. doi: 10.1016/s0264-410x(01)00445-5

Stone-Marschat, M., Moss, S., Burrage, T., Barber, M., Roy, P., & Laegreid, W. (1996). Immunization with VP2 is sufficient for protection against lethal challenge with African horse sickness virus Type 4. *Virology*, *220*(1), 219-222. doi: 10.1006/viro.1996.0304

van Rijn, P. A., Maris-Veldhuis, M. A., Potgieter, C. A., & Van Gennip, R. G. (2018). African horse sickness virus (AHSV) with a deletion of 77 amino acids in NS3/NS3a protein is not virulent and a safe promising AHS Disabled Infectious Single Animal (DISA) vaccine platform. *Vaccine*, *36*(15), 1925-1933. doi: 10.1016/j.vaccine.2018.03.003

Van Der Saag, M., Ward, M., & Kirkland, P. (2017). Application of an embryonated chicken egg model to assess the vector competence of Australian Culicoides midges for bluetongue viruses. *Med Vet Entomol*, *31*(3), 263–271. doi: 10.1111/mve.12231

Venter, G., Graham, S., & Hamblin, C. (2000). African horse sickness epidemiology: vector competence of South African Culicoides species for virus serotypes 3, 5 and 8. *Med Vet Entomol*, *14*(3), 245–250. doi: 10.1046/j.1365-2915.2000.00245.x

Venter, G., Groenewald, D., Paweska, J., Venter, E., & Howell, P. (1999). Vector competence of selected south african culicoides species for the bryanston serotype of equine encephalosis virus. *Med Vet Entomol*, *13*(4), 393–400. doi: 10.1046/j.1365-2915.1999.00188.x

Venter, G., & Paweska, J. (2007). Virus recovery rates for wild-type and live-attenuated vaccine strains of African horse sickness virus serotype 7 in orally infected South African Culicoides species. *Med Vet Entomol*, *21*(4), 377–383. doi: 10.1111/j.1365-2915.2007.00706.x

Venter, G., Wright, I., & Paweska, J. (2010). A comparison of the susceptibility of the biting midge culicoides imicola to infection with recent and historical isolates of african horse sickness virus. *Med Vet Entomol*, *24*(3), 324–328. doi: 10.1111/j.1365-2915.2010.00895.x

Venter, G., Wright, I., Van Der Linde, T., & Paweska, J. (2009). The oral susceptibility of South African field populations of Culicoides to African horse sickness virus. *Med Vet Entomol*, *23*(4), 367–378. doi: 10.1111/j.1365-2915.2009.00829.x

von Teichman, B. F., Dungu, B., & Smit, T. K. (2010). In vivo cross-protection to African horse sickness Serotypes 5 and 9 after vaccination with Serotypes 8 and 6. *Vaccine*, *28*(39), 6505-6517. doi: 10.1016/j.vaccine.2010.06.105

Whitworth, S. (1930). Memorandum on horse-sickness immunization. In *Papers veterinary section, no. 30.* Union of South Africa, Dept. of Agriculture; Pretoria: University of Pretoria, Dept. of Library Services (Digital publisher). Retrieved from `https://repository.up.ac.za/bitstream/handle/2263/13834/30_paper30_whitworth.pdf?sequence=1` (Accessed: 2021-09-28)

Wittmann, E., Mellor, P., & Baylis, M. (2002). Effect of temperature on the transmission of orbiviruses by the biting midge, *Culicoides sonorensis*. *Med Vet Entomol*, *16*(2), 147-156. doi: 10.1046/j.1365-2915.2002.00357.x

APPENDIX F

Inference for a spatio-temporal model with partial spatial data: African horse sickness virus in Morocco: Supplementary material

Inference for a spatio-temporal model with partial spatial data: African horse sickness virus in Morocco

Emma L. Fairbanks[1], Matthew Baylis[2], Janet M. Daly[1], and Michael J. Tildesley[3]

[1]School of Veterinary Medicine and Science, University of Nottingham, Loughborough. LE12 5RD, UK
[2]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Leahurst Campus, Neston, Cheshire, CH64 7TE, UK
[3]The Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research, School of Life Sciences and Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

Supplementary file 1: Kernel selection

A PubMed search conducted on 31 July 2020 using the search terms (disease OR virus) AND ∼model AND kernel AND ∼transmission' found 88 articles. Of these, 47 were excluded during a first screening of their titles for not involving modelling livestock or the type of model used and one study was repeated within the search. After reading the remaining 41 articles, a further 29 were excluded using the same criteria or for not estimating kernel parameters leaving 12 eligible articles (Table S1). The process of the systematic search is described by the PRISMA flow diagram in Figure S1.

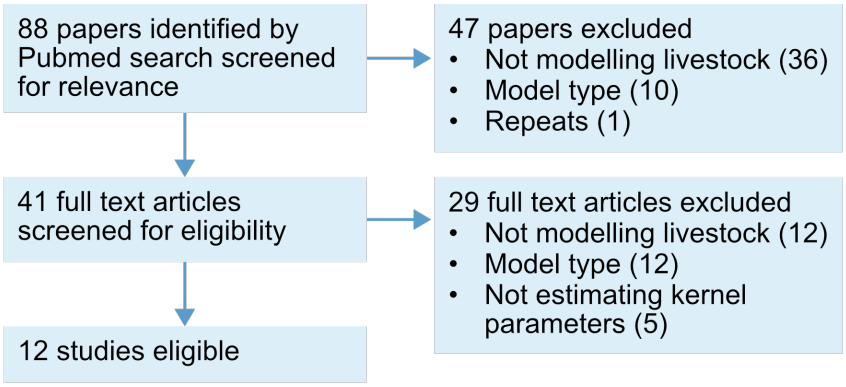

Figure S1: PRISMA flow diagram describing the process of the systematic review [10].

ID	Function	Virus	Studies	Ref
K1	$\frac{1}{1+(d/d_0)^{\alpha}}$	Foot and mouth	Hayama *et al.* (2013)	[9]
			Schnell *et al.* (2019)	[14]
		Bluetongue	Boender *et al.* (2014)	[2]
		Avian influenza	Bonney *et al.* (2018)	[4]
			Boender *et al.* (2007)a	[3]
			Boender *et al.* (2007)b	[1]
			Dorigatti *et al.* (2010)	[7]
		Swine fever	Gamado *et al.* (2017)	[5]
K2	$1 - exp(-(d/d_0)^{-\alpha})$	Foot and mouth	Hayama *et al.* (2013)	[9]
		Avian influenza	Bonney *et al.* (2018)	[4]
		Swine fever	Gamado *et al.* (2017)	[5]
K3	$exp(-(d/d_0)^{\alpha})$	Foot and mouth	Hayama *et al.* (2013)	[9]
		Avian influenza	Bonney *et al.* (2018)	[4]
K4	$(1 + d/d_0)^{-\alpha}$	Foot and mouth	Hayama *et al.* (2013)	[9]
		Avian influenza	Bonney *et al.* (2018)	[4]
		Swine fever	Gamado *et al.* (2017)	[5]
K5	$\alpha exp(-\alpha d)$	Bluetongue	Szmaragd *et al.* (2009)	[15]
		Swine fever	Gamado *et al.* (2017)	[5]
K6	$\frac{\alpha}{\pi} exp(-\alpha^2 d^2)$	Bluetongue	Szmaragd *et al.* (2009)	[15]
K7	$\frac{\alpha}{4} exp(-\alpha^{1/2} d^{1/2})$	Bluetongue	Szmaragd *et al.* (2009)	[15]

K8	$(d/d_0)^{-\alpha}$	Foot and mouth	Tildesley *at al.* (2012)*	[16]
		Avian influenza	Rorres *at al.* (2011)a	[12]
			Rorres *at al.* (2011)b	[13]
K9	$\begin{cases}(1-(d/d_0)^2)^2, & \text{if } d \leq d_0.\\ 0, & \text{otherwise.}\end{cases}$	Avian influenza	Boender *et al.* (2007)a	[3]
K10	$\begin{cases}\frac{1}{2}(1+\epsilon), & \text{if } d \leq \frac{d_0}{3}.\\ \frac{1}{2}, & \text{if } \frac{d_0}{3} < d \leq \frac{2d_0}{3}.\\ \frac{1}{2}-\frac{1}{10}\epsilon, & \text{if } \frac{2d_0}{3} < d \leq d_0.\\ 0, & \text{otherwise.}\end{cases}$	Avian influenza	Boender *et al.* (2007)a	[3]
K11	$exp(-(d/d_0 - x)^2)$	Avian influenza	Boender *et al.* (2007)a	[3]
K12	$\frac{1}{1+r^\alpha}$	Avian influenza	Boender *et al.* (2007)b	[1]

Table S1: Qualitative synthesis of the studies found fitting parameters for a distance-based kernel model of disease spread between livestock premises. * In Tildesley et al. [16] α is set to -1 [16].

Kernel K1 was the most commonly used kernel. During initial attempts to fit data to this kernel parameter estimates did not converge or yield reproducible results. Therefore, we considered kernels K5–K7 as these were used to model bluetongue virus, spread by the same insects as AHSV. Szmaragd et al. [15] found that the kernel K6 performed better while attempting to model the spread of BTV-8 in Northern Europe during 2006, therefore we attempted to fit this kernel. These kernels could all be written with the same formula as

$$K = exp(-\alpha d^\epsilon),$$

where α is the coefficient multiplied by d in the index of the exponential and the coefficient multiplied by the exponential is accounted for by the parameter β in the force of infection (Equation 1). Szmaragd *et al.* (2009) found that K6, when ϵ was set to 2, provided the best fit for the data.

Supplementary file 2: Parameter estimation

2.1 Sequential Monte Carlo-Markov chain Monte Carlo (SMC-MCMC)

As results in supplementary file 3 show there was little variability in spatial homogeneity across the randomly generated spatial distributions we perform 10 SMC iterations each MCMC iteration. Each SMC iteration we generate a new parameter set and spatial distribution. After the initial iteration of LHS parameters, the parameter sets that generated the three highest log-likelihoods (30%) were accepted. We then create two new variables; qu denoting the lowest log-likelihood accepted and $best_{loglik}$ denoting the largest log-likelihood for all the parameter sets. In future iterations, parameter sets that produce log-likelihoods larger than qu are accepted. For the 1000 main MCMC iterations we then have three possible cases: no parameter sets are accepted, one parameter set is accepted and two or three parameters sets are accepted.

Case 1: Two or three parameter sets accepted.
If the largest log-likelihood is larger than $best_{loglik}$ then: $best_{params}$ is set to the parameter set which generated this log-likelihood and $best_{loglik}$ is set to this likelihood. qu is then updated to the lowest log-likelihood accepted in the current iteration. σ is set to the standard deviation between the parameters accepted the current iteration and σ_{max} is set to 3σ.

Case 2: One parameter set accepted.
If the accepted log-likelihood is larger the $best_{loglik}$ then: $best_{params}$ is set to the parameter set which generated this log-likelihood and $best_{loglik}$ is set to this likelihood. qu is not updated. σ is updated to the minimum of 1.1σ from the previous iteration and σ_{max}.

Case 3: No parameter sets accepted
qu is not updated. σ is updated to the minimum of 1.1σ from the previous iteration and σ_{max}.

Parameters from the next iteration are then drawn from a normal distribution with mean $best_{params}$ and standard deviation 1.1σ with a random noise between $\pm 10\%$ of $best_{params}$.

2.2 Reducing computational cost of the likelihood function

In order to reduce the computational cost of calculating the log-likelihood, the contribution for each time-step ($LogLik_t$) was calculated individually. Initially, this was set to 0 for each time-step. The days were sorted into four cases: the first day, days where there was an infection the previous day, days where there was an infection the current day but not the previous and days where there was no infection on either the previous or current day.

Case 1: The first day.
The first infection is considered to have occurred on day 0. Therefore, to calculate $LogLik_1$ we consider the force of infection (Equation 1) on all premises not infected on day 0 by those infected on day 0. $LogLik_1$ is then calculated as the summation of $ln(1 - e^{-\lambda})$ for premises infected on day 1 and $-\lambda$ for uninfected premises.

Case 2: Days where there was an infection the previous day.
The force of infection for each premises can be calculated as

$$\lambda_t = \lambda_{t-1} + \lambda_{\text{from premises infected on day } t-1}.$$

$LogLik_t$ is then calculated as the summation of $ln(1 - e^{-\lambda_t})$ for premises infected on day t and $-\lambda_t$ for uninfected premises.

Case 3: Days with a new infection where there was no new infection the previous day.
$LogLik_t = LogLik_{t-1}$ for uninfected premises. For premises infected on day t the force of infection (λ) remains the same and

$$\begin{aligned} LogLik_t =& LogLik_{t-1} - \text{probability premises was not infected} \\ &+ \text{probability premises was infected}, \\ =& LogLik_{t-1} - ln(1 - e^{-\lambda}) + -\lambda. \end{aligned}$$

Case 4: Days where there was no infection of the previous or current day.
There is no change in the force of infection and therefore $LogLik_t = LogLik_{t-1}$.

The total log-likelihood is the summation of the contributions of each time-step.

Supplementary file 3: Homogeneity of spatial distributions

3.1 Methods

A homogeneous spatial distribution is where in any circle in a given area you will find approximately the same number of points. If a spatial distribution is not homogeneous it may be clustered to a certain spatial scale. Here we use Ripley's L function to detect deviations from spatial homogeneity and compare these between generated spatial distributions for each spatial specificity.

In order to calculate Ripley's L function we first need to calculate Ripley's K function defined as

$$K(t) = \frac{A}{N} \sum_{i \neq j} \frac{I(d_{ij} < t)}{N},$$

where A is the area of the plane containing all points, N is the number of points, d_{ij} is the Euclidean distance between the i^{th} and j^{th} point, t is the search radius and I is a Boolean function which is 1 when its operand is and 0 otherwise [6]. The points are approximately homogeneous when $K \approx \pi t^2$. From here we can calculate Ripley's L function (the variance stabilised Ripley's K function) defined as

$$L(t) = \left(\frac{K(t)}{\pi} \right)^{1/2},$$

which for spatial homogeneous data has expected value t with constant variance.

We use Ripley's L function to analyse the difference in spatial distributions for the three spatial specificities used in this study (village, region, province). For each spatial specificity, we calculate Ripley's L function for 50 random spatial distributions for t = [0 60]. Calculating Ripley's K and L function for a data-set this size would be computationally expensive. Therefore, we suggest a modified version of the Ripley's K function given as

$$\hat{K}(t) = \frac{A}{N_j} \sum_{i \neq j} \frac{I(d_{ij} < t)}{N_i}$$

where j is the set of all premises, i is a subset of j containing infected premises. N_j and N_i are the total number of premises and the number of infected premises, respectively.

To calculate the area of the shapefiles the Matlab function *deg2utm* was used to convert latitude and longitude co-ordinates given in the shapefiles into Universal Transverse Mercator (UTM) coordinates [11]. UTM co-ordinates give a zone number and easting-northing pair. As the points are in the same UTM zone we assume the Euclidean distance between them can be approximated as

$$\sqrt{(e1 - e2)^2 + (n1 - n2)^2},$$

for a given easting-northing pair (e1,n1) and (e2,n2). The area of the polygon was then calculated using the Matlab function *polyarea*.

The distance between premises was calculated using the haversine function for the great-circle distance (the shortest distance between the points). Here for two sets of co-ordinates ($lat1$, $log1$) and ($lat2$, $lon2$) we first convert to radians giving

$$\begin{aligned}\phi_1 =& \frac{lat1\pi}{180},\\ \phi_2 =& \frac{lat2\pi}{180},\\ \lambda_1 =& \frac{lon1\pi}{180} \text{ and}\\ \lambda_2 =& \frac{lon2\pi}{180}.\end{aligned}$$

Then:

$$a = sin^2(\frac{\phi_2 - \phi_1}{2}) + cos(\phi_1)cos(\phi_2)sin^2(\frac{\lambda_2 - \lambda_1}{2,})$$

and

$$2atan2(\sqrt{a}, \sqrt{1-a}).$$

Using this we can calculate the distance between two points in km as

$$d = Rc,$$

where R is the radius of the earth (6371 km).

2.2 Results

We find that the spatial distributions are approximately spatially homogeneous for shorter distances from infected premises (<15km). For distances larger than approximately 15km, Ripley's L function increases implying a homogeneous population. However, for each spatial specificity and iteration this is consistent with little variability (Figure S2).

A reason for this deviation from spatial homogeneity for larger distances could be 'the edge problem' in the Ripley's K algorithm. The algorithm assumes an infinitely continuous planar space, which when applied to geographical data sets is not a reality. Here, the borders of the shapefile (for example coasts) will affect the output value. Various approaches have been suggested to correct for these boundary issues [8, 17]. However, here we are more concerned with the consistency between iterations, and little variability is observed. Therefore, we conclude that a large number of different spatial distributions is not necessary at each MCMC iteration.

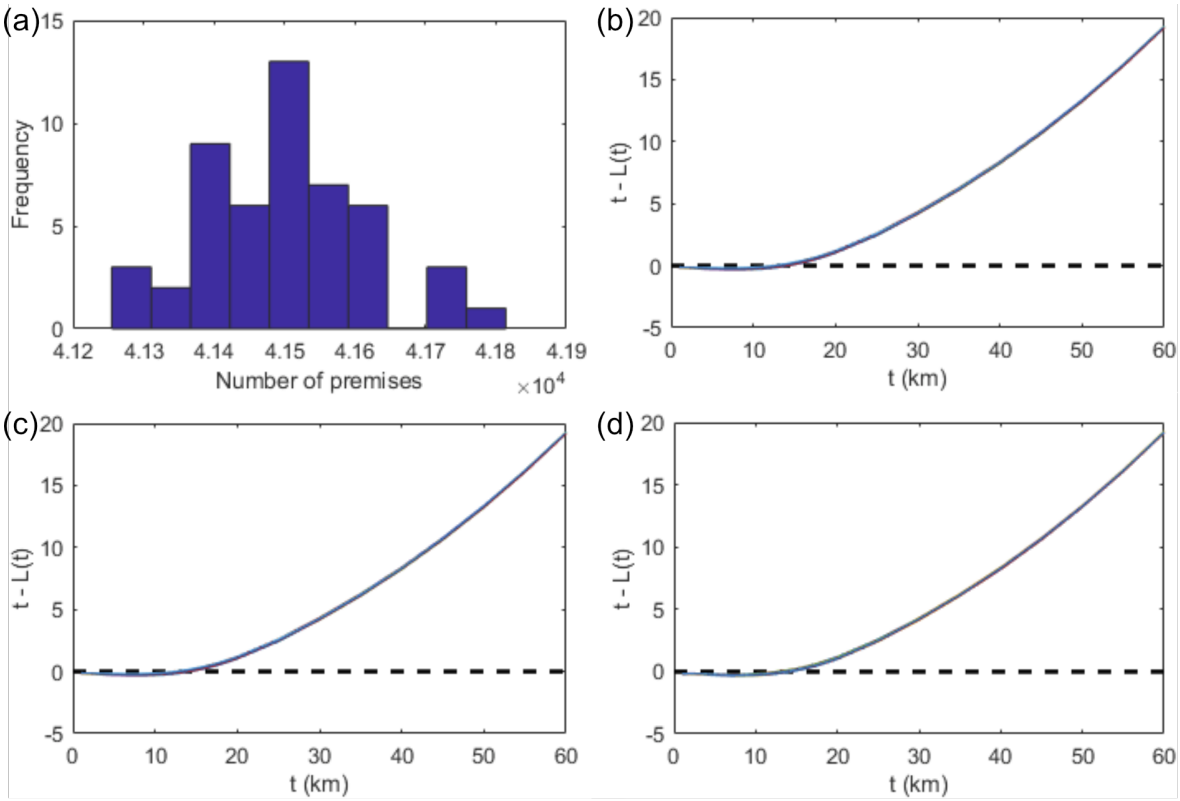

Figure S2: Ripley's L function for the three spatial specificities. (a) The number of premises each iteration for 50 random spatial distributions generated from the latitude and longitude data. (b),(c) and (d) 50 iterations of Ripley's L function generated from the latitude and longitude (village), region, and province data, respectively.

Supplementary figure S3

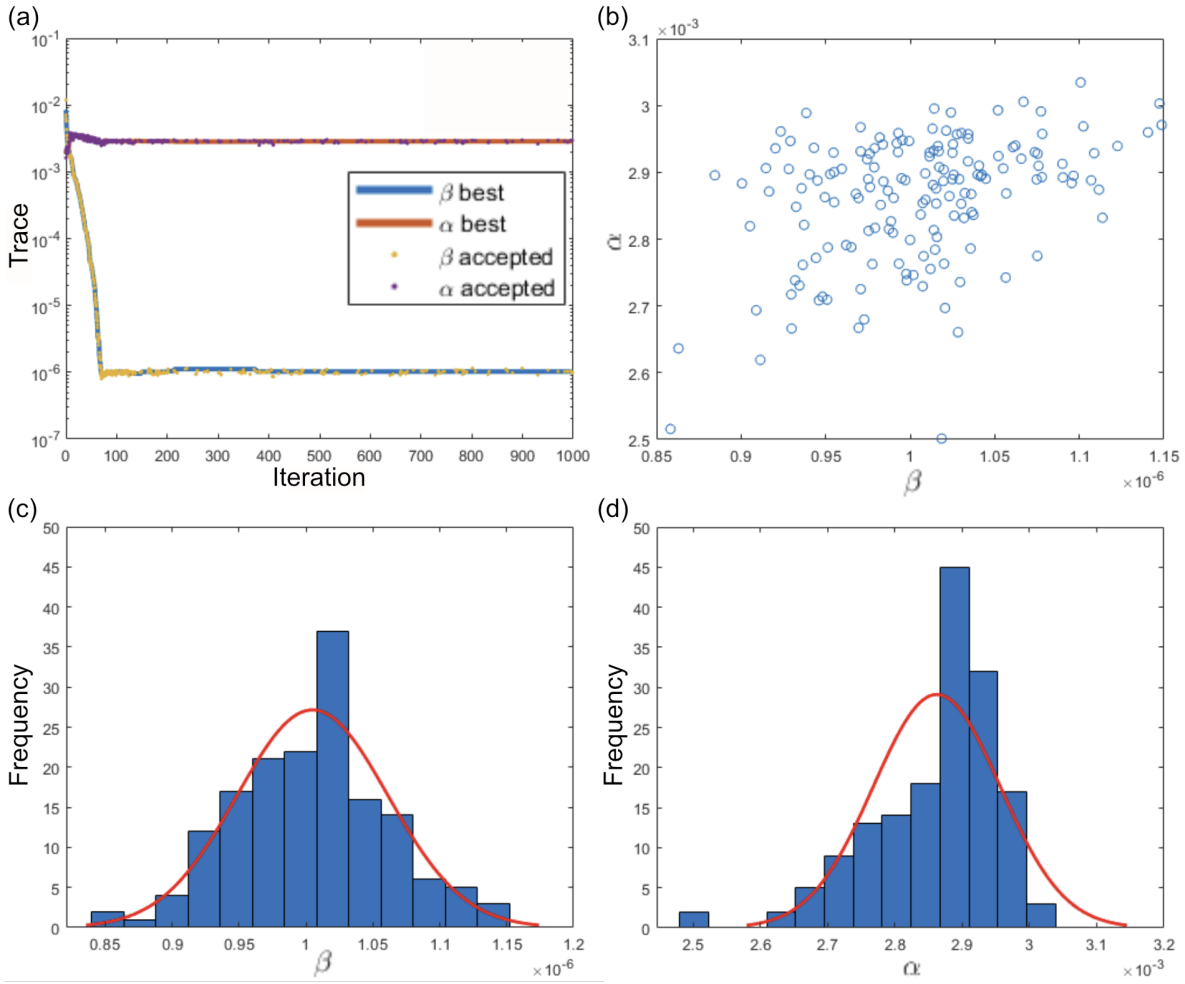

Figure S3: Results from the parameter estimation using the latitude and longitude village data. (a) The trace of the parameters which generated the largest log-likelihood over all previous simulations(lines) and the accepted values (dots). (b) The correlation between accepted parameters once the burn-in (first 100 iterations) is removed. (c) and (d) show histograms and the probability density of a normal distribution fitted to accepted values (once burn-in removed) of β and α, respectively.

Supplementary figure S4

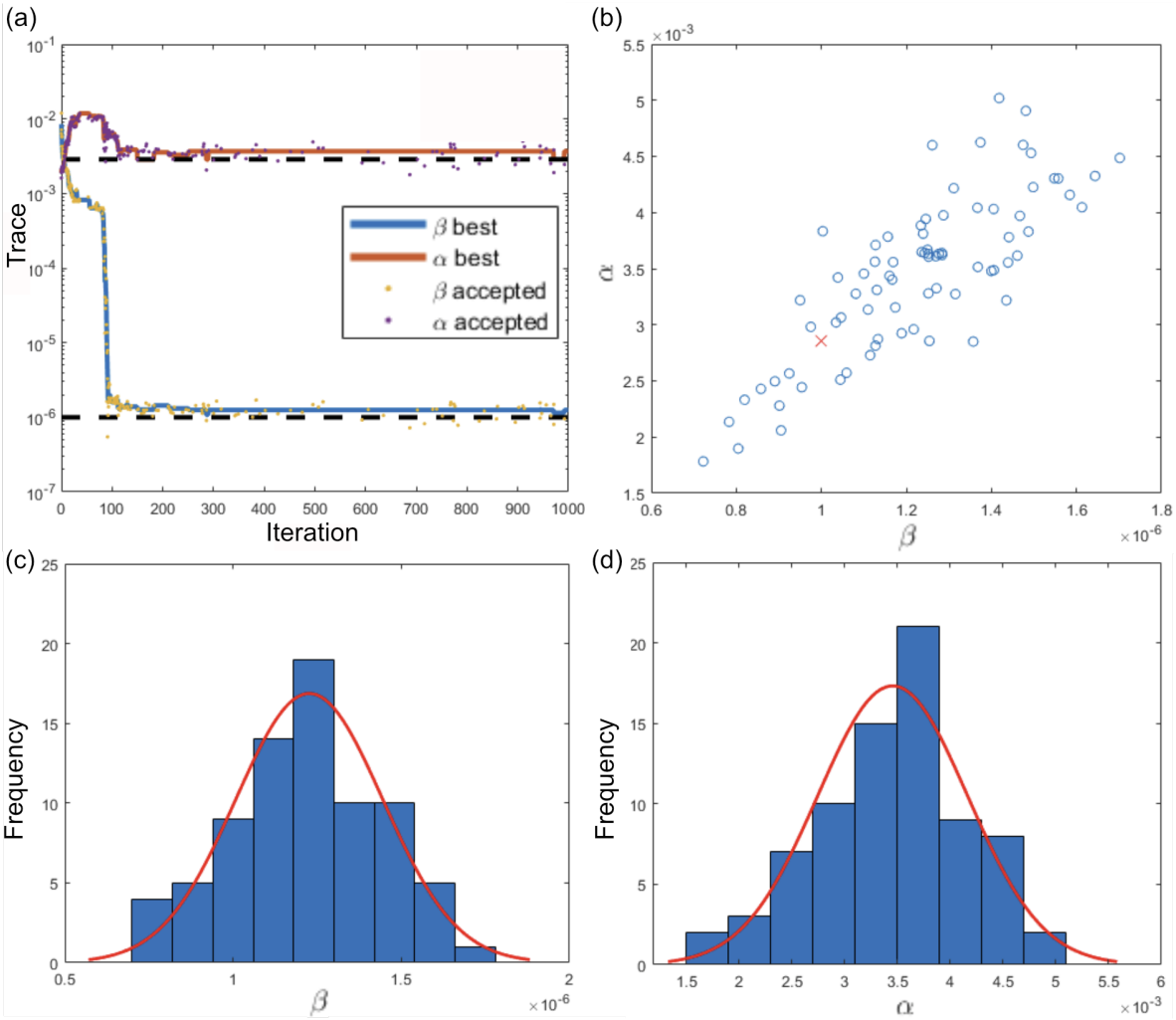

Figure S4: Results from the parameter estimation using the region data. (a) The trace of the parameters which generated the largest log-likelihood over all previous simulations(lines) and the accepted values (dots). (b) The correlation between accepted parameters once the burn-in (first 150 iterations) is removed. (c) and (d) show histograms and the probability density of a normal distribution fitted to accepted values (once burn-in removed) of β and α, respectively.

Supplementary figure S5

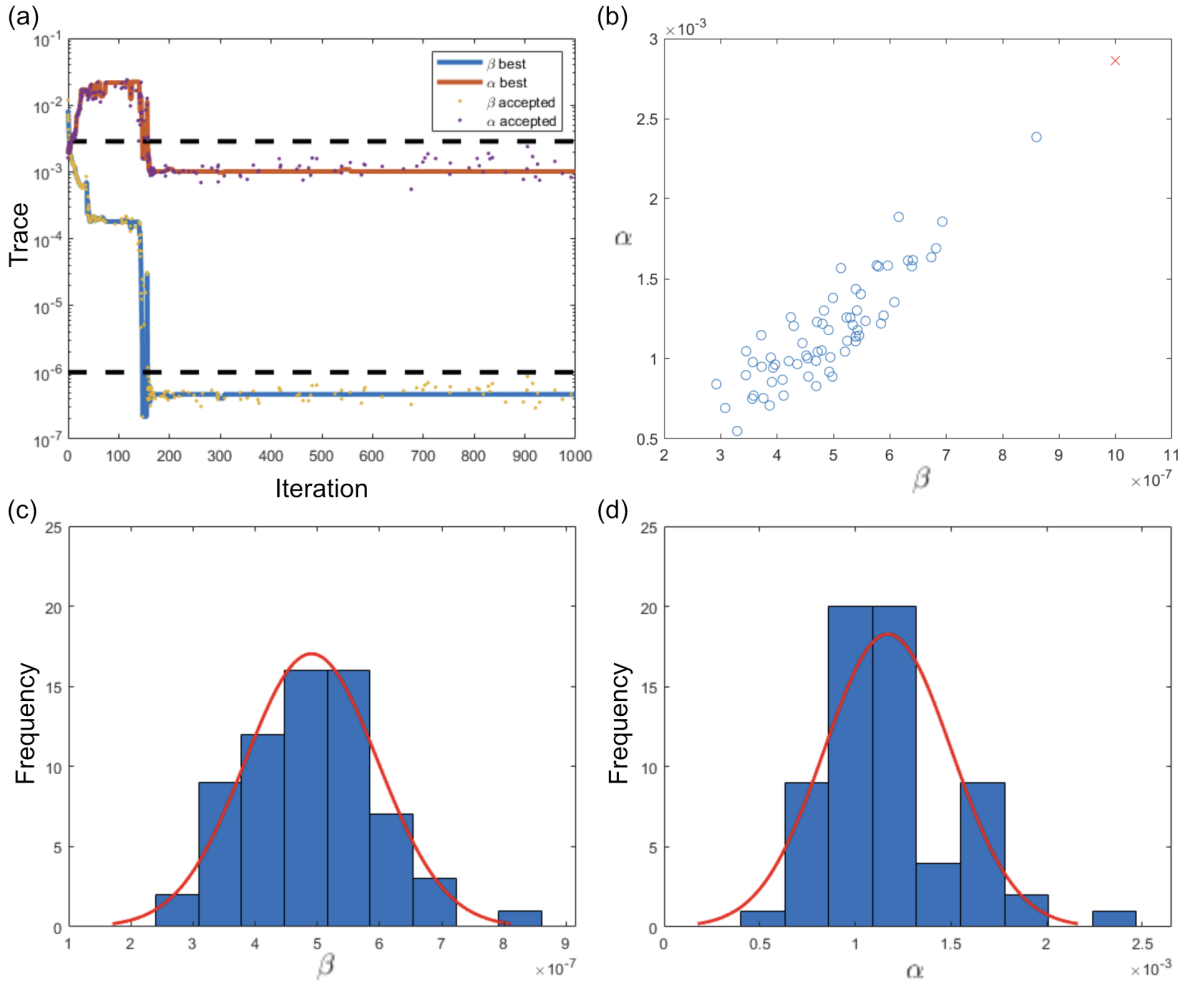

Figure S5: Results from the parameter estimation using the province data. (a) The trace of the parameters which generated the largest log-likelihood over all previous simulations(lines) and the accepted values (dots). (b) The correlation between accepted parameters once the burn-in (first 200 iterations) is removed. (c) and (d) show histograms and the probability density of a normal distribution fitted to accepted values (once burn-in removed) of β and α, respectively.

References

[1] Boender, G. J., Hagenaars, T. J., Bouma, A., Nodelijk, Elbers, A. R., de Jong, M. C. M. and Van Boven, M. [2007], 'Risk maps for the spread of highly pathogenic avian influenza in poultry', *PLoS Comput Biol* **3**(4), e71.

[2] Boender, G. J., Hagenaars, T. J., Elbers, A. R., Gethmann, J. M., Meroc, E., G, H. and De Koeijer, A. A. [2014], 'Confirmation of spatial patterns and temperature effects in bluetongue virus serotype-8 transmission in NW-Europe from the 2007 reported case data', *Vet Res* **45**(1), 75.

[3] Boender, G. J., Meester, R., Gies, E. and De Jong, M. C. M. [2007], 'The local threshold for geographical spread of infectious diseases between farms', *Prev Vet Med* **82**(1-2), 90–101.

[4] Bonney, P. J., Malladi, S., Boender, G. J., Weaver, J. T., Ssematimba, A., Halvorson, D. A. and Cardona, C. J. [2018], 'Spatial transmission of H5N2 highly pathogenic avian influenza between Minnesota poultry premises during the 2015 outbreak', *PloS one* **13**(9), e0204262.

[5] *Data-Driven Risk (Assessment from Small Scale Epidemics: Estimation and Model Choice for Spatio-Temporal Data with Application to a Classical Swine Fever O* [n.d.].

[6] Dixon, P., El-Shaarawi, A. H. and Piegorsch, W. W. [2012], 'Ripley's K function', *Encyclopedia of Environmetrics* **3**.

[7] Dorigatti, I., Mulatti, P., Rosà, R., Pugliese, A. and Busani, L. [2010], 'Modelling the spatial spread of H7N1 avian influenza virus among poultry farms in Italy', *Epidemics* **2**(1), 29–35.

[8] Ge, Y., Sun, M. and Pu, Y. [2019], 'Geographic information system-based edge effect correction for Ripley's K-function under irregular boundaries', *Geogr Res* **57**(4), 436–447.

[9] Hayama, Y., Yamamoto, T., Kobayashi, S., Muroga, N. and Tsutsui, T. [2013], 'Mathematical model of the 2010 foot-and-mouth disease epidemic in Japan and evaluation of control measures', *Prev Vet Med* **112**(3-4), 183–193.

[10] Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J. and Moher, D. [2009], 'The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration', *J Clin Epidemiol* **62**(10), e1–e34.

[11] *Rafael Palacios* [2020], https://www.mathworks.com/matlabcentral/fileexchange/10915-deg2utm. deg2utm - File Exchange - MATLAB Central. Accessed: 2020-08-24.

[12] Rorres, C., Pelletier, S. T. K., Bruhn, M. C. and Smith, G. [2011], 'Ongoing estimation of the epidemic parameters of a stochastic, spatial, discrete-time model for a 1983–84 avian influenza epidemic', *Avian Dis* **55**(1), 35–42.

[13] Rorres, C., Pelletier, S. T. K. and Smith, G. [2011], 'Stochastic modeling of animal epidemics using data collected over three different spatial scales', *Epidemics* **3**(2), 61–70.

[14] Schnell, P. M., Shao, Y., Pomeroy, L. W., Tien, J. H., Moritz, M. and Garabed, R. [2019], 'Modeling the role of carrier and mobile herds on foot-and-mouth disease virus endemicity in the Far North Region of Cameroon', *Epidemics* **29**, 100355.

[15] Szmaragd, C., Wilson, A. J., Carpenter, S., Wood, J. L. N., Mellor, P. S. and Gubbins, S. [2009], 'A modeling framework to describe the transmission of bluetongue virus within and between farms in Great Britain', *PloS one* **4**(11), e7741.

[16] Tildesley, M. J., Smith, G. and Keeling, M. J. [2012], 'Modeling the spread and control of foot-and-mouth disease in Pennsylvania following its discovery and options for control', *Prev Vet Med* **104**(3-4), 224–239.

[17] Yamada, I. and Rogerson, P. A. [2003], 'An empirical comparison of edge effect correction methods applied to K-function analysis', *Geographical analysis* **35**(2), 97–109.

Appendix G

Influence of setting-dependent contacts and protective behaviours on asymptomatic SARS-CoV-2 infection amongst members of a UK university: Supplementary material

Influence of setting-dependent contacts and protective behaviours on asymptomatic SARS-CoV-2 infection amongst members of a UK university

Emma Fairbanks[1,2], Kirsty Bolton[2,*], Ru Jia[3], Grazziela Figueredo[4], Holly Knight[3], and Kavita Vedhara[3]

[1]School of Veterinary Medicine and Science, University of Nottingham
[2]School of Mathematical Sciences, University of Nottingham
[3]School of Medicine, University of Nottingham
[4]School of Computer Science, University of Nottingham
[*]Corresponding author: kirsty.bolton@nottingham.ac.uk

Supplementary file 1: Data collection and curation

A summary of participant test and survey dates by test outcome and participant age is given in Figure S1.

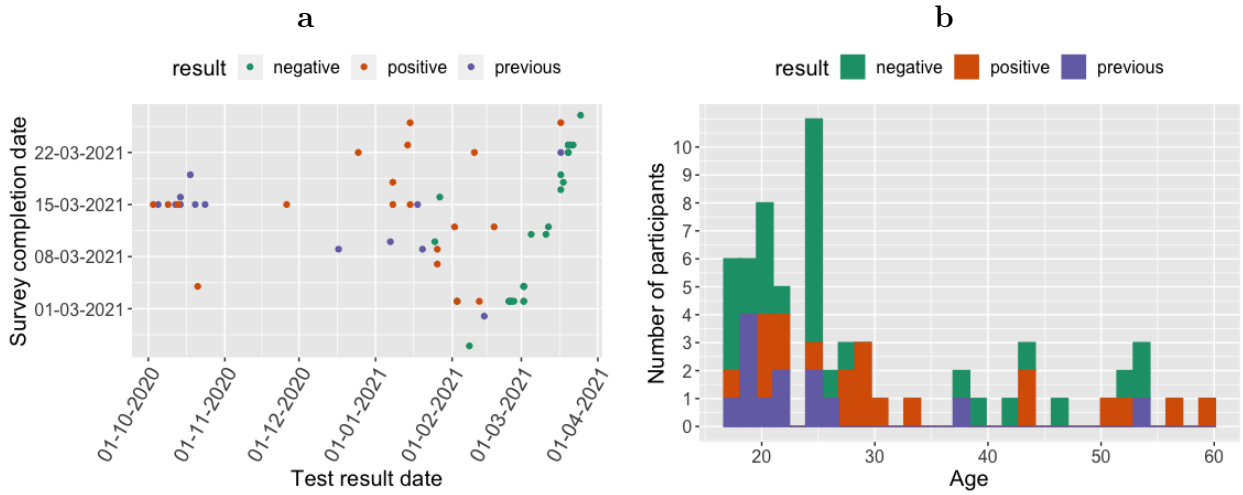

Figure S1: (a) Date of receipt of SARS-CoV-2 PCR test result versus date of survey completion. (b) Age and asymptomatic SARS-CoV-2 PCR test result of participants.

Activities involving laboratory work — including those undertaken by postgraduate students — and the activities of laboratory technicians, were classed as occurring in the research setting. Postgraduates activities indicating use of an office or activities involving one-to-one laboratory training were also classified as research. Activities were classed as occurring in a teaching setting if they involved students or postgraduate students/staff members teaching/supporting the session. Activities in this category included practicals, lectures and placements. All other activities on campus were categorised as occurring in the "Campus other" setting. Notable amongst these was one participant working in a role involving student interactions and two participants who accessed the library. Often participants would provide information in the "additional comments" field that enabled their activities to be classified by setting. However, motivation for this appeared to reduce over the course of the survey. Therefore, if an individual described their university work as qualifying for the research setting at any point in the questionnaire it was assumed that all their work fell into that setting category. In some cases participants did not offer any additional details and therefore activities could not be assigned to the research or teaching setting. These participant's campus activities were categorised as occurring in "Campus other".

The setting non-private travel was used to describe all travel which was not a personal car or air-travel, these included public transport and taxis. The setting abroad/aeroplane is used to describe settings which include both travel by aeroplane and holiday activities. Notable here in an individual whose activities included skiing in Italy and spa activities. Initial analysis showed that including these activities in other relevant settings (for example skiing in exercise) heavily influenced outputs. Some participants many have been abroad and but

their activities or additional comments did not flag this.

Exercise may have been in a fitness centre or outdoors and included walking. Retail included essential and non-essential retail, which may have been click and collect-like or drive-through services. Non-university work describes activities should as working in a pub, supermarket and as a carer. In our sample these activities were only undertaken by students. The social setting captures activities which were not performed in another setting (such as at another household or recreation ground). Testing describes the participant taking a SARS-CoV-2 test. All other activities were classified as "other".

For one activity, in the "Campus other" setting, a participant stated that they visited the library for 360 minutes and had 360 contacts. The university regulations did not permit this large number of individuals in library spaces at this time. Therefore, we assume this was a mistake while filling out the questionnaire. We set the number of contacts for this activity to 3, consistent with other activities reported in this setting.

Supplementary table S2: Summary statistics for setting-specific contact measures

Co-variate	Measure	Participated	#activities	Contacts	Duration	PCH	Contacts*	PCH*
HH size	mean	3.02						
	sd	2.23						
Abroad/	mean	0.06	0.20	6.87	0.34	3.34	6.52	2.62
aeroplane	sd	0.24	1.15	27.51	2.37	23.39	25.92	18.32
Campus other	mean	0.14	0.39	2.40	2.09	15.141	2.40	15.14
	sd	0.35	1.04	9.77	6.31	73.95	9.77	73.95
Exercise	mean	0.51	1.47	4.88	1.87	8.22	3.71	6.59
	sd	0.51	2.31	15.29	3.67	27.71	14.08	26.31
Hospitality	mean	0.12	0.24	4.33	0.32	8.69	3.96	8.19
	sd	0.33	0.85	19.36	1.12	34.45	17.70	32.53
Non-university	mean	0.06	0.14	4.44	0.71	41.89	4.38	41.15
work	sd	0.24	0.74	20.07	2.92	203.96	19.78	199.72
Non-private	mean	0.18	0.53	9.04	0.30	6.70	8.10	6.26
travel	sd	0.39	1.56	37.54	0.81	26.87	34.12	25.17
Other	mean	0.29	0.59	3.88	0.53	3.18	2.51	2.09
	sd	0.46	1.50	12.62	1.57	12.73	9.88	11.11
Research	mean	0.24	1.29	6.38	4.79	26.26	5.95	24.92
	sd	0.43	2.62	16.95	10.55	75.20	16.83	75.04
Retail	mean	0.59	1.04	41.33	0.49	22.69	39.90	21.95
	sd	0.50	1.38	56.47	0.85	36.14	54.27	34.12
Social	mean	0.31	0.47	1.61	0.67	3.13	0.98	2.36
	sd	0.47	0.87	4.21	1.78	12.07	4.00	11.84
Teaching	mean	0.22	0.55	9.08	1.57	28.33	6.38	19.50
	sd	0.42	1.60	27.10	5.93	102.74	18.65	69.10
Testing	mean	0.33	0.39	1.52	0.12	0.29	1.36	0.22
	sd	0.47	0.64	3.06	0.34	0.62	2.94	0.50

Table S1: The mean and standard deviation between participants of the binary measure of whether the participant participated in an activity, the number of distinct activities, total contacts , total duration of activities, PCH, the total non-household contacts and non-household PCH in each setting.

Supplementary file 2: Comparison of model performance for setting-specific contact measures based on different contact definitions

Predictor	Median	Minimum	Maximum	PMP
Participated	-126.7505	-126.76	-126.74	0.39
Activities	-127.44	-127.45	-127.43	0.20
Contacts	-129.20	-129.20	-129.19	0.03
Duration	-128.25	-128.26	-128.24	0.09
PCH	-128.26	-128.27	-128.26	0.08
Contacts*	-129.36	-129.37	-129.35	0.03
PCH*	-127.54	-127.55	-127.51	0.18

Table S2: The median, minimum and maximum of the marginal log-likelihood estimates for the multi-variate logistic model for each predictor, as well as the mean posterior model probabilities (PMPs). * = predictors were only non-household contacts are considered. Results are given to 2 decimal places.

Household size, campus other, hospitality, retail and testing do not have supported coefficient values all with the same sign for any of the predictors. Non-university work has only positive coefficients for all contact measure definitions apart from contacts and non-household contacts, which were the models which received the smallest PMP values. These contact definitions also supported non-negative values of coefficients for the teaching covariate, additional to PCH (which received the smallest PMP after contacts and non-household contacts). Research did not have all-negative supported coefficients with the predictors number of activities and non-household contacts. However, in all these cases the distribution is heavily skewed favouring values with the sign supported in all other models.

Contact definition	Mean/median (standard deviation) Positive	Negative	Previous
Participated	0.60/1.00 (0.49)	0.04/0.00 (0.18)	0.23/0.00 (0.42)
Activities	0.70/1.00 (0.46)	0.00/0.00 (0.01)	0.46/0.00 (0.50)
PCH*	0.05/0.00 (0.22)	0.00/0.00 (0.03)	0.16/0.00 (0.36)
Duration	0.50/0.40 (0.50)	0.07/0.00 (0.07)	0.31/0.00 (0.46)
PCH	0.31/0.00 (0.46)	0.00/0.00 (0.01)	0.23/0.00 (0.42)
Contacts	0.00/0.40 (0.49)	0.00/0.03 (0.17)	0.00/0.08 (0.27)
Contacts*	0.42/0.00 (0.49)	0.03/0.00 (0.18)	0.08/0.03 (0.27)

Table S3: Predicted test outcomes for logistic regression based on different sets of contact measures.

Supplemntary file 3: Validation of Bayesian logistic regression

The median bounds of SIs from the leave-one-out deletions have the same sign as the analysis including all participants for all contact definitions and settings, apart from the support intervals (SIs) for the coefficient of non-household contacts in the social setting those for PCH in the teaching setting. These exceptions correspond to model contact definitions with low PMPs (0.03 and 0.08 respectively). Furthermore, the variation in SIs when performing the leave-one-out deletion model fits is consistent with the conclusions drawn from the full analysis. In the full analysis, non-household contact is the only contact definition to have only negative values in the SI for the social setting. Similarly, the leave-one-out deletion yields median SI bounds for the PCH contact definition that are consistent with models with larger PMPs (with all values in the SI being negative). In some cases analysis on all participants did not yield an SI, however the leave-one-out deletion analysis did.

Contact description	Lower bound Value*	Median	sd	Upper bound Value*	Median	sd
Constant						
Participated	-8.12	-8.12	0.86	4.63	4.67	0.58
Activities	-8.38	-8.53	0.20	4.51	4.72	0.20
Contacts	-8.03	-8.39	0.49	4.81	4.68	1.75
Duration	-8.54	-8.44	0.51	4.68	4.66	0.47
PCH	-8.58	-8.49	1.04	4.68	4.69	0.69
Contacts*	-8.39	-8.39	1.22	4.76	4.70	0.48
PCH*	-6.87	-8.62	0.56	3.87	4.73	0.59
HH size						
Participated	-0.85	-0.87	0.08	0.35	0.35	0.08
Activities	-0.90	-0.89	0.07	0.30	0.33	0.07

Contacts	-0.70	-0.78	0.07	0.46	0.45	0.06
Duration	-0.92	-0.91	0.08	0.31	0.32	0.07
PCH	-0.65	-0.67	0.07	0.51	0.51	0.07
Contacts*	-0.71	-0.78	0.07	0.45	0.45	0.06
PCH*	-0.65	-0.64	0.067	0.52	0.53	0.07
Holiday/aeroplane						
Participated	1.91	1.95	0.22	3.19	3.04	0.31
Activities	-	1.09	1.46	-	2.22	1.45
Contacts	-	0.56	1.95	-	1.11	2.54
Duration	-	1.26	0.00	-	1.39	0.00
PCH	-	3.68	0.29	-	4.94	1.65
Contacts*	-	0.49	1.78	-	1.03	4.04
PCH*	-	6.02	1.36	-	16.46	17.22
Campus other						
Participated	-0.68	-0.68	0.13	0.57	0.58	0.13
Activities	-0.35	-0.41	0.10	0.83	0.83	0.08
Contacts	-0.27	-0.26	0.08	0.92	0.93	0.09
Duration	-0.27	-0.26	0.08	0.91	0.91	0.08
PCH	-0.11	-0.11	0.06	1.09	1.10	0.08
Contacts*	-0.21	-0.20	0.09	0.98	0.99	0.09
PCH*	-0.12	-0.11	0.07	1.08	1.13	0.09
Exercise						
Participated	-1.22	-1.23	0.07	0.04	0.06	0.07
Activities	-2.00	-2.10	0.12	-0.51	-0.51	0.07
Contacts	-0.29	-0.28	0.14	1.00	1.02	0.16
Duration	-1.61	-1.66	0.07	-0.25	-0.24	0.03
PCH	-0.95	-0.96	0.10	0.31	0.25	0.11
Contacts*	-0.23	-0.25	0.12	1.05	1.04	0.14
PCH*	-0.74	-0.76	0.22	0.42	0.41	0.12
Hospitality						
Participated	-0.92	-0.91	0.18	0.26	0.27	0.13
Activities	-0.28	-0.29	0.06	0.81	0.81	0.06
Contacts	-	-	-	-	-	-
Duration	-0.03	-0.02	0.08	1.25	1.24	0.10
PCH	-	-	-	-	-	-
Contacts*	-	-	-	-	-	-
PCH*	-	-13.49	3.28	-	-5.78	0.16
Non-university work						
Participated	2.14	2.12	0.10	5.38	5.32	0.31
Activities	0.77	0.78	0.10	2.45	2.39	0.18
Contacts	-0.02	-0.02	0.07	1.18	1.23	0.11
Duration	0.21	0.21	0.05	1.50	1.52	0.06
PCH	0.20	0.11	0.60	1.04	1.10	0.95
Contacts*	-0.03	-0.01	0.05	1.18	1.26	0.11

PCH*	0.25	0.24	0.66	1.13	1.10	2.66
Non-private travel						
Participated	-1.26	-1.27	0.13	0.01	0.02	0.09
Activities	-1.25	-1.17	0.09	0.10	0.10	0.06
Contacts	-1.20	-1.11	0.09	0.06	0.06	0.05
Duration	-1.06	-1.07	0.11	0.25	0.26	0.09
PCH	-1.10	-1.13	0.10	-0.24	-0.24	0.07
Contacts*	-1.13	0.04	-1.13	0.07	0.04	0.03
PCH*	-1.09	-1.10	0.09	-0.23	-0.21	0.07
Other						
Participated	-0.54	-0.48	0.09	0.74	0.74	0.09
Activities	-0.89	-0.89	0.09	0.32	0.32	0.07
Contacts	-1.29	-1.32	0.07	-0.05	-0.04	0.05
Duration	-1.15	-1.12	0.16	0.16	0.16	0.07
PCH	-	-0.81	0.00	-	0.44	0.00
Contacts*	-1.10	-1.07	0.05	0.08	0.09	0.03
PCH*	-	-20.21	26.72	-	1.83	11.32
Research						
Participated	-1.80	-1.82	0.10	-0.32	-0.32	0.08
Activities	-1.37	-1.34	0.08	0.03	0.04	0.07
Contacts	-1.28	-1.35	0.06	-0.05	-0.04	0.03
Duration	-1.31	-1.32	0.06	-0.05	-0.04	0.04
PCH	-1.32	-1.34	0.04	-0.06	-0.06	0.02
Contacts*	-1.19	-1.20	0.05	0.03	0.02	0.04
PCH*	-1.25	-1.21	0.05	-0.00	-0.00	0.03
Retail						
Participated	-0.37	-0.35	0.06	0.94	0.94	0.07
Activities	-1.18	-1.23	0.07	0.04	0.05	0.05
Contacts	-0.89	-0.91	0.06	0.25	0.26	0.06
Duration	-0.89	-0.89	0.09	0.30	0.30	0.09
PCH	-0.74	-0.73	0.07	0.46	0.46	0.07
Contacts*	-1.07	-1.04	0.07	0.19	0.19	0.05
PCH*	-0.66	-0.71	0.07	0.48	0.49	0.06
Social						
Participated	-0.35	-0.36	0.05	0.91	0.92	0.06
Activities	-0.04	-0.06	0.07	1.29	1.30	0.07
Contacts	-0.91	-0.92	0.11	0.25	0.18	0.08
Duration	-0.05	0.02†	0.10	1.31	1.36	0.11
PCH	0.24	0.24	0.32	1.17	1.25	0.34
Contacts*	-1.05	-1.03	0.09	-0.01	0.03†	0.08
PCH*	-	0.79	1.21	-	1.28	2.62
Teaching						
Participated	-2.45	-2.46	0.10	-0.78	-0.78	0.05

Activities	-1.70	-1.71	0.11	-0.33	-0.29	0.06
Contacts	-1.35	-1.34	0.04	0.02	0.01	0.04
Duration	-1.38	-1.25	0.08	-0.04	-0.11	0.06
PCH	-1.34	-1.33	0.10	0.06	-0.03†	0.06
Contacts*	-1.22	-1.24	0.06	0.03	0.02	0.04
PCH*	-1.41	-1.51	0.32	-0.41	-0.52	0.31
Testing						
Participated	-0.96	-0.96	0.11	0.36	0.36	0.09
Activities	-0.64	-0.59	0.09	0.67	0.68	0.08
Contacts	-0.53	-0.51	0.07	0.71	0.72	0.07
Duration	-0.82	-0.82	0.05	0.43	0.43	0.06
PCH	-0.72	-0.69	0.08	0.51	0.50	0.07
Contacts*	-0.33	-0.36	0.06	0.87	0.87	0.08
PCH*	-0.61	-0.59	0.08	0.61	0.56	0.08

Table S4: Median and standard deviation of results from leave-one-out deletion for the upper and lower bounds of SIs for contact measures. * Value corresponds to the bound of the SI when all data is included in the analysis, as in Table 2. † Sign of result different to when all data is included in the analysis.

Supplementary file 4: Protective behaviours

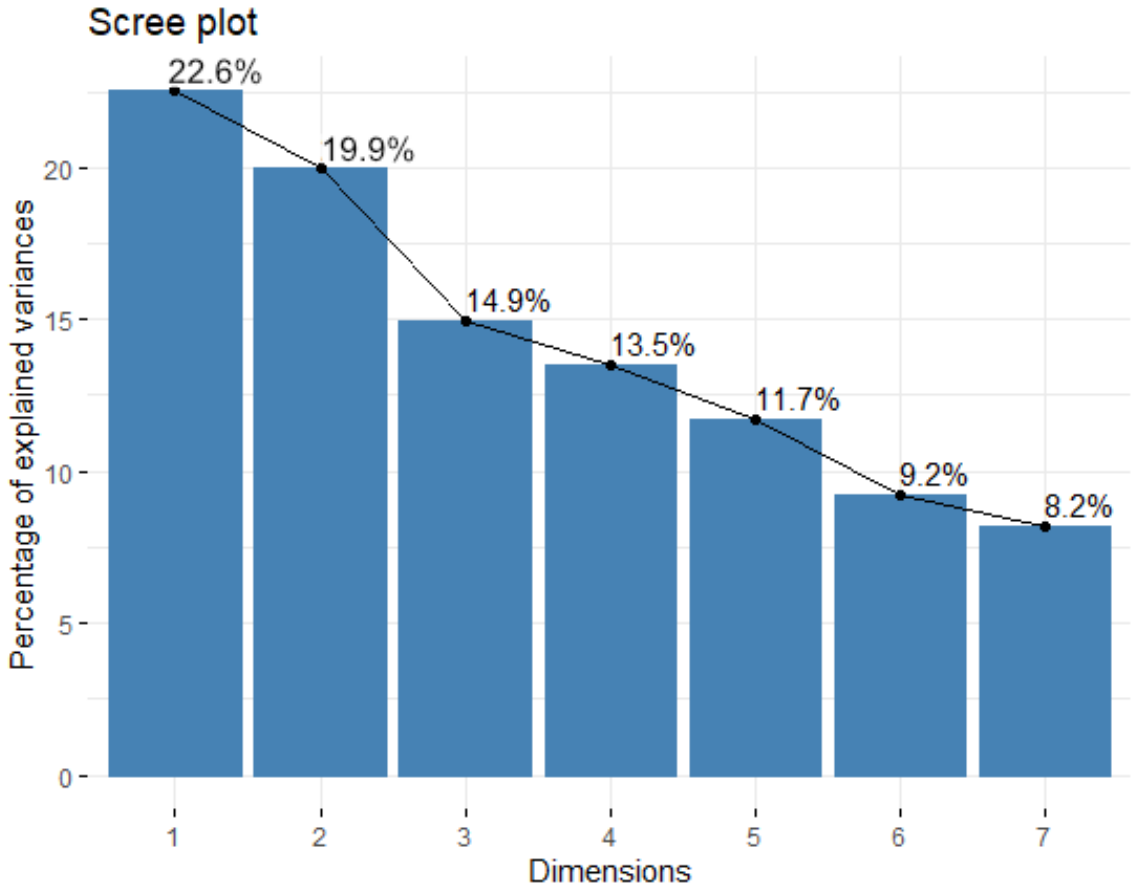

Figure S2: Variance described by each dimension in the MCA analysis.

Variable	Dim 1 (22.5%)		Dim 2 (20.0%)		Dim 3 (14.9%)		Dim 4 (13.5%)	
Quantitative	Correl.	p	Correl.	p	Correl.	p	Correl.	p
Age*	0.14	7.2e-3	0.13	9.9e-3	_	_	_	_
Qualitative	R^2	p	R^2	p	R^2	p	R^2	p
Gender*	_	_	_	_	_	_	0.02	1.5e-2
Role*	0.04	1.0e-3	_	_	_	_	_	_
Result*	0.04	1.1e-4	0.04	6.4e-5	_	_	_	_
Setting*	0.46	9.8e-41	0.34	4.7e-26	0.17	7.1e-10	0.15	5.6e-8
Environment	0.42	4.5e41	0.36	1.6e-34	0.11	6.2e-09	0.56	1.9e-62
Mask	0.51	2.1e-55	0.04	8.4e-4	_	_	0.03	3.2e-3
SD	0.09	1.3e-7	0.57	3.3e-66	_	_	_	_
Hands	0.58	7.6e-65	0.46	1.3e-45	0.93	1.7e-197	0.36	8.2e-33

Table S5: MCA results. * supplementary variables (not included in dimension calculation). Function used to calculate = r dimdesc. 4 dimensions describe 70% of variance - each dimension is given in the brackets.

For the Fisher's tests categorical properties were converted into Boolean measures, coded 1 or 0 if property was true or false respectively, for each activity. The true options for the Boolean measures were whether an activity was associated with a female, UG, PG, staff, positive test result, each setting (1 variable for each), indoors, ventilated (either outdoors or ventilated indoors), mask worn, socially distanced at all times, washed hands before, washed hands after and washed hands both before and after. Some properties could have more than two answers (role, setting, environment and hand washing) and are split into multiple Boolean measures.

Correlated variables		OR	95% OR CI		p value
Female	UG	6.472795	3.493324	12.71164	4.36E-10
Female	Staff	0.337524	0.21185	0.533905	1.13E-05
Female	Positive	2.664691	1.599047	4.529741	0.000469
Female	Abroad/aeroplane	Inf	1.530765	Inf	0.02225
Staff	Abroad/aeroplane	12.2377	1.664423	540.6127	0.011299
Positive	Abroad/aeroplane	Inf	4.976757	Inf	8.71E-05
PG	Campus other	0	0	0.573969	0.018168
Staff	Campus other	3.870203	1.280474	14.05116	0.025649

Positive	Hospitality	11.24038	2.337989	107.3193	0.001599
Positive	Non-university work	Inf	3.120526	Inf	0.001599
Staff	Other	3.758366	1.543411	10.11401	0.005749
Female	Research	0.30956	0.16708	0.562325	0.000254
UG	Research	0	0	0.107677	6.22E-10
PG	Research	7.039633	3.786352	13.33705	1.09E-09
Positive	Research	0.214044	0.079453	0.493209	0.000217
UG	Teaching	12.43208	4.413549	43.29909	2.76E-07
Staff	Teaching	0.043895	0.001063	0.273694	2.85E-05
Positive	Teaching	0.152793	0.017266	0.63121	0.009242
PG	Ventilated	0.4818	0.282386	0.822384	0.018302
Exercise	Ventilated	51.47862	8.662316	2073.156	8.39E-12
Hospitality	Ventilated	0.156996	0.026785	0.646021	0.01201
Non-university work	Ventilated	0.080149	0.001728	0.673413	0.021085
Research	Ventilated	0.304279	0.164898	0.555779	0.000301
Retail	Ventilated	0.380887	0.19778	0.729087	0.00783
Social	Ventilated	5.696474	1.351154	51.01485	0.029157
PG	Indoors	2.038997	1.161187	3.674717	0.029157
Exercise	Indoors	0.034354	0.0126	0.079868	5.81E-23
Hospitality	Indoors	Inf	1.781015	Inf	0.015058
Research	Indoors	Inf	13.94208	Inf	2.23E-13
Retail	Indoors	2.238987	1.091414	4.906826	0.049792
Social	Indoors	0.113769	0.027468	0.354014	7.98E-05
Teaching	Indoors	8.225179	1.983399	72.93845	0.002343
Positive	Mask	0.513287	0.309268	0.85161	0.021085
Abroad/aeroplane	Mask	0	0	0.336928	0.001918
Exercise	Mask	0.232761	0.128445	0.415815	2.77E-06
Hospitality	Mask	0.093413	0.009776	0.449727	0.002052
Travel	Mask	14.02407	2.232678	582	0.001612
Research	Mask	4.802925	2.073896	12.98889	0.000186
Retail	Mask	12.77424	3.211127	111.1286	1.91E-05
Social	Mask	0.022366	0.000536	0.144744	1.15E-07
Teaching	Mask	Inf	2.823971	Inf	0.000961
Indoors	Mask	4.082006	2.441941	6.89965	2.22E-07
PG	SD	2.000447	1.164936	3.501119	0.025873
Positive	SD	0.441388	0.269018	0.720046	0.00265
Abroad/aeroplane	SD	0	0	0.504648	0.009137
Exercise	SD	3.504704	1.844847	7.01555	0.000202
Travel	SD	0.304539	0.111036	0.763133	0.021085
Retail	SD	0.420856	0.205431	0.841077	0.026294
Ventilated	SD	2.606503	1.577373	4.337191	0.000592
Indoors	SD	0.54059	0.327	0.885405	0.029403
PG	No hands	0.052152	0.001277	0.315869	7.70E-05

Staff	No hands	3.081542	1.539377	6.415828	0.002903
Positive	No hands	3.916108	1.976742	7.937215	0.000232
Abroad/aeroplane	No hands	72.79709	9.626162	3206.739	8.78E-07
Travel	No hands	10.62188	4.1694	27.59858	2.77E-06
Research	No hands	0.089935	0.002189	0.550464	0.005749
Retail	No hands	0.114598	0.002779	0.706546	0.021145
Indoors	No hands	0.345876	0.156526	0.741046	0.01201
Mask	No hands	0.261738	0.126618	0.527625	0.00037
SD	No hands	0.108521	0.039392	0.25783	5.02E-08
PG	Hands before	2.127534	1.170659	4.027228	0.025873
Abroad/aeroplane	Hands before	0.044193	0.001001	0.326248	0.000633
Campus other	Hands before	7.733918	1.180849	326.9385	0.04522
Exercise	Hands before	0.490605	0.275869	0.877285	0.03556
Travel	Hands before	0.336782	0.136612	0.817318	0.035196
Research	Hands before	4.038419	1.746327	10.90549	0.001122
Teaching	Hands before	Inf	3.061419	Inf	0.000469
Indoors	Hands before	3.1059	1.852618	5.249085	5.98E-05
Mask	Hands before	3.538563	2.097997	6.011858	6.70E-06
Female	Hands after	0.437693	0.205932	0.877194	0.039494
PG	Hands after	7.117743	2.206516	36.59253	0.000446
Staff	Hands after	0.373618	0.191725	0.709843	0.005749
Positive	Hands after	0.252148	0.13008	0.479725	8.72E-05
Abroad/aeroplane	Hands after	0.016663	0.000378	0.125221	2.77E-06
Travel	Hands after	0.117277	0.045819	0.294809	1.67E-05
Research	Hands after	13.22285	2.173833	541.5632	0.001557
Mask	Hands after	3.494713	1.816303	6.835463	0.000465
SD	Hands after	8.446876	3.855998	20.55385	1.16E-08
PG	Hands both	2.028788	1.137624	3.74398	0.037134
Abroad/aeroplane	Hands both	0.04874	0.001104	0.359476	0.00107
Campus other	Hands both	8.507178	1.300038	359.4156	0.044095
Travel	Hands both	0.373344	0.151746	0.904635	0.044095
Research	Hands both	4.48648	1.943084	12.10406	0.000455
Teaching	Hands both	12.97079	2.067061	538.1547	0.002903
Indoors	Hands both	2.697677	1.629032	4.493621	0.000341
Mask	Hands both	3.53807	2.114156	5.964829	5.62E-06
SD	Hands both	1.887969	1.154352	3.099289	0.026142

Table S6: Significant Fisher's test results - Benjamini-Hochberg adjusted p-values. PG = Postgraduate, UG = undergraduate, Positive = positive test result, Mask = face covering worn, SD = socially distanced, No hands = no hand washing, Hands before/after/both = hands washed before/after/both.