Epidemiological inference at the threshold of data availability: an influenza A(H1N2)v spillover event in the United Kingdom

John A. Fozard[1], Emma C. Thomson[1,2], Christopher J. R. Illingworth[1,*]

[1]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

[2]Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, United Kingdom

*Corresponding Author: christopher.illingworth@glasgow.ac.uk

*ORCID ID: 0000-0002-0030-2784

Abstract

Viruses which infect animals regularly spill over into the human population, but individual events may lead to anything from a single case to a novel pandemic. Rapidly gaining an understanding of a spillover event is critical to calibrating a public health response. We here propose a novel method, using likelihood free rejection sampling to evaluate the properties of an outbreak of swine-origin influenza A(H1N2)v in the United Kingdom, detected in November 2023. From the limited data available we generate historical estimates of the probability that the outbreak had died out in the days following the detection of the first case. Our method suggests that the outbreak could have been said to be over with 95% certainty between 19 and 29 days after the first case was detected, depending upon the probability of a case being detected. We further estimate the number of undetected cases conditional upon the outbreak still being live, the epidemiological parameter R_0, and the date on which the spillover event itself occurred. Our method requires minimal data to be effective. While our calculations were performed after the

event, the real-time application of our method has potential value for public health responses to cases of emerging viral infection.

Introduction

Viral transmission from animals to humans poses a serious threat in terms of its potential to generate novel pandemics[1]. Influenza viruses have a track record of causing serious impacts upon human health, with the 1918, 1957, and 1968 pandemics each being associated with more than one million deaths[2,3].

While such pandemics are rare events, they occur in a context of much more frequent animal-to-human spillover events: In 2022, epidemiological monitoring detected close to 60 cases of avian and swine influenza infection in humans worldwide[4]. While most of these events do not lead to large numbers of people being infected, at the very earliest stages of detection the distinction between an outbreak that will remain localised, and an outbreak that will go on to seed a pandemic, may be small. Efforts are required to understand spillover events in their earliest stages.

Statistical approaches for understanding outbreaks work on very different quantities of data. The SARS-CoV-2 pandemic saw the implementation of a broad range of computational and statistical approaches to track the nature and impact of the virus. Studies early in the pandemic characterised the epidemiological properties of the virus[5,6], and produced estimates of the epidemiological reproductive number R_0 in different contexts[7]. Methods for 'nowcasting' combined multiple datasets to estimate local levels of viral prevalence[8,9]. A UK-wide project generated and shared hundreds of thousands of SARS-CoV-2 viral sequences[10]. Genome sequence data were used to study virus evolution and transmission on multiple scales[11–13].

Soon after a spillover occurs, data limitations may present themselves. Where multiple instances of infection from an emerging pathogen are observed, epidemiological models can highlight possible viral adaptation [14]. Inferences can be made of the epidemiological transmission parameter R_0[15], and of differences in R_0 across settings[16]. Sequence data can be used to assess the evolutionary origins and epidemiological characteristics of emergent viruses[17,18].

At the very earliest stages of an outbreak, data limitations may be severe; the detection of a spillover event may begin with a single case of infection. To explore what might be achieved in this minimal case, we here investigate a case of human influenza A(H1N2)v virus infection, observed the UK in November 2023. The detected case had no known contact with pigs, so was here assumed not to be the first case in the outbreak[19]. After the one detection, no further observations of infection were made. Considering the period immediately following the detection, we use a novel method to estimate the probability of the outbreak having ceased, and the potential number of undetected cases of infection. We discuss the potential for minimal datasets to inform the public health response in the days following the detection of a viral spillover event.

Results

As a precursor to our method, we estimated the probability of a single case of influenza A(H1N2)v being detected as being between 4% and 10%. Publicly available data shows that in week 48 of 2023 the UK rate of GP consultations for influenza-like illness (ILI) was 4.6 per 100,000 individuals, equal to a national total of approximately 3,100 consultations per week. Following these consultations, 557 samples were tested as part of a sentinel swabbing scheme,

suggesting that approximately 18% of GP consultations for ILI led to swabbing and further testing[20]. Following the detection of a case, swabbing was increased in the Yorkshire and Humber region, where the case was found[20]. While data specific to the UK population are sparse, published estimates suggest that between 25 and 50% of individuals with ILI might seek healthcare[21,22]. Our estimate range was derived from these values.

We evaluated the properties of the A(H1N2)v influenza outbreak using likelihood free rejection sampling, first carrying out a form of historical nowcasting. Given a day following the detection of the outbreak, and supposing the knowledge that no further cases had been detected up until that day, we generated large numbers of simulated influenza outbreaks, identifying the statistical properties of simulations that matched the observed data up until the considered day. In this manner we estimated that, fourteen days after the first detection of A(H1N2)v influenza, the probability that the outbreak had died out was between 66% and 88% (Figure 1A). The range in this value reflects uncertainty in the probability of a case being detected, with higher detection probabilities leading to greater probabilities of the outbreak having ended. At a detection rate of 4%, we inferred that the outbreak could be said to have ended with 95% confidence by day 29 after the first detection; the same conclusion could be drawn by day 19 given a detection rate of 10%.

We further estimated the number of active cases of infection, conditional upon the outbreak having not died out. In the considered case, the median number of active cases fourteen days after the detection of the first case was estimated as between 55 and 91 (Figure 1B). Estimated numbers of active cases decreased slowly with time, considering days successively further from that of the detection (Supplementary Figure 1). Our model therefore suggested a dichotomy of potential circumstances: The outbreak was likely to have died out, but could have involved numbers of active cases if it was still live. We note that, soon after the detection of the first

104 case, there exists the potential for the outbreak to involve very large numbers of cases. This
105 result is explained by the 18-day delay between symptom onset and detection for the first
106 detected case[23]. Given a high R_0, this delay would have been sufficient for the virus to have
107 spread substantially within the population. More rapid identification of outbreaks limits the
108 potential for spread prior to detection.
109
110 We made a retrospective estimate of the time of the first, undetected, case of infection;
111 estimates were made a nominal date 90 days after the first detection. Our model predicted that
112 this first infection occurred 22 days prior to the first detection, on 1st November, only a few days
113 before the first detected case became symptomatic (Figure 1B). The inferred likelihood function
114 is steeply skewed, with the mean of this distribution between 25 and 28 days prior to the first
115 detection, varying with the detection rate. However, the interval between the time of the first
116 infection and the time at which the first detected individual became symptomatic was likely
117 short, with the detected individual potentially being a direct contact of the index case.
118
119 We also made a retrospective estimate of R_0 for this outbreak, obtaining a value of 0.9 (Figure
120 1D). This lies below the threshold necessary to sustain an outbreak, and is lower than
121 estimates for seasonal influenza viruses, for which R_0 is in the region of 1.3[24]. The very
122 sparse data used in making our estimate led to a large degree of uncertainty in this estimate;
123 further data would sharpen the inferred distribution.
124
125 Exploring our data further, we examined the extent to which our retrospective estimates of R_0
126 and the time of first infection could have been made closer to the observation of the first case.
127 In the first few days after the detection of the case there was little power to rule out large values
128 of R_0 (Supplementary Figure 2A), but large values were progressively excluded with time. By
129 contrast the inferred time of the first case of infection was relatively stable, with estimates

calculated a few days after the observation being very close to our final estimate (Supplementary Figure 2B).

Our results are dependent upon the prior distribution chosen for R_0. By default we used a uniform prior between zero and 4, but the highest values in this range are beyond those previously observed even for the 1918 pandemic virus[25]. Recalculating results with a uniform prior between zero and 2 did not strongly affect our results (Supplementary Figure 3).

Discussion

Examining data describing a spillover event of a swine influenza virus into the human population we used a method of rejection sampling to explore what can be learnt, both at the time of the immediate response, and in retrospect, from the limited public data describing this event. Our method provided time-dependent estimates of the probability that the outbreak had died out, and for the number of undetected cases in the case that the outbreak was still live. It further provided estimates for the date of the spillover event itself, and for the parameter R_0. Our model achieves what it does because both the observation of a case of infection, and the subsequent non-observation of cases, are informative for the model: The failure to observe a second case of infection is an important piece of data.

Our work pushes at the boundary of the amount of data required to learn about a spillover event, showing that even minimal data are sufficient to draw early and provisional conclusions about the outbreak. Our approach may be of value in a public health context: Estimating the certainty with which we can say whether an outbreak is likely to have died out could inform decisions about the investigation of cases and the resources applied to this task. Simultaneously, the provisional nature of our conclusions should not be underemphasised.

Methods such as contact tracing would provide substantially more information about the potential for the virus to have spread, while genome sequencing of any further cases would facilitate phylogenetic and other genomic approaches to epidemiology.

Our approach is limited by its simplicity. For example, we assumed a homogeneous population, in which each infected person is equally infectious to the others, and neglected the potential for evolution to change the infectivity of the virus. We modelled data collection in a simple manner, assuming for example a fixed time between symptom onset and test result. We note that the epidemiological dynamics of infection, expressed as distributions of the time to symptom onset and to infecting others, are of critical importance to our method, but were of necessity based upon distributions inferred for other influenza strains: The A/H1N2 virus detected potentially would not mirror seasonal influenza in this way. Many of the assumptions made by our method could be elaborated upon, for example by allowing for heterogeneity in transmission[26]. The limited data available in this case did not encourage the use of more complex models.

The variation in our results under different detection scenarios highlights the potential value of improved systems for detecting cases of infection following a spillover event. In this case, detection efforts were stepped up in the region of the detection of the first case. More thorough and faster testing provides a greater certainty that a spillover has not led to a persistent outbreak and reduces the potential for an outbreak to grow undetected.

While we have applied our model to a specific event, describing the spillover of influenza A(H1N2)v into the human population in the UK, our approach has the potential for broader use. Given reasonable estimates for epidemiological parameters, viruses other than influenza could also be modelled. Events involving more than one detection of a positive case could also be assessed, though as the number of cases of infection increases our approach becomes less

182 computationally efficient. Once large amounts of data become available, alternative methods
183 for epidemiological inference are likely to perform better than our own. Our approach is of
184 potential value in the first stages of an outbreak when data are most limited.
185
186 **Methods**
187
188 We used likelihood free rejection sampling to evaluate the likely state of the outbreak underlying
189 the available data. This method generated a large number of simulated outbreaks, retaining
190 only those which were compatible with the data collected from the A(H1N2)v outbreak, counting
191 the number of cases detected on each day, and assuming that a given number of days, denoted
192 by t_o, had passed since the day on which the first detected case was observed. For the
193 historical now-casting calculations simulations which exactly matched the data up to time t_o
194 were accepted. The R_0 parameters of the accepted simulations provided samples of the
195 posterior distribution of this statistic. The properties of these simulations for different values of t_o
196 were used to estimate the properties of the outbreak. Where retrospective estimates of
197 parameters were made, these were calculated at the point t_o = 90 days.
198
199 *Simulation of outbreaks*
200 Time was modelled discretely, in units of whole days. Each simulated outbreak started with a
201 primary case. The simulation then proceeded day by day. Infected individuals were assumed
202 to become symptomatic a random number of days t_s after being infected. Every symptomatic
203 individual then infected a total of R others, with the time of each infection event occurring a
204 random number of days t_i after symptom onset, and where R was Poisson distributed with
205 parameter R_0. The random numbers t_s and t_i were drawn from Weibull distributions with
206 parameters based upon published parameters describing influenza infection[27], but with

samples rounded to the nearest integer. Specifically, where w is the cumulative distribution function

$$w(x, a, b) = 1 - e^{-\left(\frac{x}{b}\right)^a}$$

t_s had probability mass function

$$P(\mathrm{t_s} = t) = w\left(t_s + \frac{1}{2}, \mathrm{a}, \mathrm{b}\right) - w\left(\max\left\{t_s - \frac{1}{2}, 0\right\}, \mathrm{a}, \mathrm{b}\right),$$

And similarly for t_i. Parameters for the distribution of t_s were given by a=7.4026 and b=1.7375, while parameters for the distribution of t_i were given by a=1.0314 and b=1.0025.

For each infected individual in our simulation we randomly determined whether the case was detected. Detection was modelled as occurring with fixed probability p_d. Detection was assumed to occur 18 days after the day of symptom onset, following data from the influenza A(H1N2)v case[23]. The exception to this rule was the primary case in the outbreak. Following information that the index case had no contact with animals, the primary case was assumed not to have been detected[19].

Each simulated outbreak was continued until it either died out, with no more cases of infection existing, or until the first day we were certain whether or not the numbers of cases detected in the simulation matched the number of detected cases in the dataset up to the observation time.

We assumed a uniform prior over the epidemiological parameter R_0 within the window [0.1, 4.0]. Rather than randomly sample, we performed a grid search, conducting 10^6 simulations for each of 40 equally spaced discrete values of R_0 from 0.1 to 4.0.

Once simulations were complete, we calculated the proportion of simulations for each R_0 that were accepted, $P_{t_o}(\text{accepted}|R_0)$. Normalising this statistic gave the posterior probability of R_0 at a given observation time t_o,

$$P_{t_o}(R_0|\text{accepted}) = \frac{P_{t_o}(\text{accepted}|R_0)\pi(R_0)}{\sum_{R_0} P_{t_o}(\text{accepted}|R_0)\,\pi(R_0)},$$

where the prior $\pi(R_0)$ is uniform. Using the accepted simulations, we estimated properties of the viral population, using the formula

$$P(Q|accepted\ at\ t_o) = \frac{\#(simulations\ accepted\ at\ t_0\ and\ Q)}{\#(simulations\ accepted\ at\ t_0)},$$

for properties Q including the number of infected individuals being k at time t_i days after the detected case, the outbreak having died out at time t_i days after the detected case, or the time of the first infection having occurred a specific number of days before the detected case.

In the calculations above, we assumed that infection lasted for seven days following infection. We note that alternative prior distributions for R_0 could be used in our calculation; we show in Supplementary Information the results of placing a lower upper bound on this statistic.

Data/Code availability

All data used for this analysis was obtained from publicly available sources. Our code is named OINK (Outbreak Inference given Negligible Knowledge) and is available from https://github.com/cjri/OINK/.

Acknowledgements

We acknowledge funding from the UK Medical Research Council (grant number MC_UU_00034/6). We thank Margaret Hosie and Antonia Ho for comments on an earlier version of this mansuscript. We thank the Pandemic Preparedness team at the MRC University of Glasgow Centre for Virus Research, and Dr Robert Goudie for discussions.

References

1. Morse SS, Mazet JA, Woolhouse M, Parrish CR, Carroll D, Karesh WB, et al. Prediction and prevention of the next pandemic zoonosis. The Lancet. 2012;380: 1956–1965. doi:10.1016/S0140-6736(12)61684-5

2. Viboud C, Simonsen L, Fuentes R, Flores J, Miller MA, Chowell G. Global Mortality Impact of the 1957–1959 Influenza Pandemic. J Infect Dis. 2016;213: 738–745. doi:10.1093/infdis/jiv534

3. Honigsbaum M. Revisiting the 1957 and 1968 influenza pandemics. The Lancet. 2020;395: 1824–1826. doi:10.1016/S0140-6736(20)31201-0

4. European Centre for Disease Prevention and Control. Zoonotic influenza: Annual Epidemiological Report for 2022. [cited 22 Jan 2024]. Available: https://www.ecdc.europa.eu/en/publications-data/zoonotic-influenza-annual-epidemiological-report-2022

5. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26: 672–675. doi:10.1038/s41591-020-0869-5

6. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med. 2020;382: 1199–1207. doi:10.1056/NEJMoa2001316

7. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. JCM. 2020;9: 967. doi:10.3390/jcm9040967

8. Birell P, Blake J, van Leeuwen E, Gent N, De Angelis, Daniela. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. Phil Trans R Soc B. 2021;376: 20200279. doi:10.1098/rstb.2021.0001

9. Nicholson G, Lehmann B, Padellini T, Pouwels KB, Jersakova R, Lomax J, et al. Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework. Nat Microbiol. 2021;7: 97–107. doi:10.1038/s41564-021-01029-0

10. Marjanovic S, Romanelli RJ, Ali G-C, Leach B, Bonsu M, Rodriguez-Rincon D, et al. COVID-19 Genomics UK (COG-UK) Consortium: Final Report. Santa Monica, CA: RAND Corporation; 2022.

11. Du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science. 2021;371: 708–712. doi:10.1126/science.abf2946

12. Hamilton WL, Tonkin-Hill G, Smith ER, Aggarwal D, Houldcroft CJ, Warne B, et al. Genomic epidemiology of COVID-19 in care homes in the east of England. eLife. 2021;10: e64618. doi:10.7554/eLife.64618

13. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. Nat Rev Microbiol. 2023;21: 361–379. doi:10.1038/s41579-023-00878-2

14. Arinaminpathy N, McLean AR. Evolution and emergence of novel human infections. Proc R Soc B. 2009;276: 3937–3943. doi:10.1098/rspb.2009.1059

15. Blumberg S, Lloyd-Smith JO. Inference of R0 and Transmission Heterogeneity from the Size Distribution of Stuttering Chains. Ferguson N, editor. PLoS Comput Biol. 2013;9: e1002993. doi:10.1371/journal.pcbi.1002993

16. Blumberg S, Funk S, Pulliam JRC. Detecting Differential Transmissibilities That Affect the Size of Self-Limited Outbreaks. Wilke CO, editor. PLoS Pathog. 2014;10: e1004452. doi:10.1371/journal.ppat.1004452

17. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009;459: 1122–1125. doi:10.1038/nature08182

18. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. Science. 2009;324: 1557–1561. doi:10.1126/science.1176062

19. Chand M, Byrne, Alex, Bennett, Allan, Stewart, Amanda B, Hayward, Andrew, Rambaut, Andrew, et al. Influenza A(H1N2)v: rapid technical assessment - GOV.UK. 11 Dec 2023 [cited 24 Jan 2024]. Available: https://www.gov.uk/government/publications/influenza-ah1n2v-technical-briefings/influenza-ah1n2v-rapid-technical-assessment

20. UK Health Security Agency. National Influenza and COVID-19 surveillance report 7 December 2023. 2023. Available: https://assets.publishing.service.gov.uk/media/6571a95933b7f2000db7201b/Weekly_Flu_and_COVID-19_report_w49.pdf

21. Ma W, Huo X, Zhou M. The healthcare seeking rate of individuals with influenza like illness: a meta-analysis. Infectious Diseases. 2018;50: 728–735. doi:10.1080/23744235.2018.1472805

22. Chawla D, Benitez A, Xu H, Whitehill V, Tadesse-Bell S, Shapiro A, et al. Predictors of Seeking Care for Influenza-Like Illness in a Novel Digital Study. Open Forum Infectious Diseases. 2023;10: ofac675. doi:10.1093/ofid/ofac675

23. World Health Organisation. Influenza A(H1N2) variant virus infection - United Kingdom of Great Britain and Northern Ireland. 2023 Dec. Available: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON496

24. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC Infect Dis. 2014;14: 480. doi:10.1186/1471-2334-14-480

25. Mills CE, Robins JM, Lipsitch M. Transmissibility of 1918 pandemic influenza. Nature. 2004;432: 904–906. doi:10.1038/nature03063

26. Wegehaupt O, Endo A, Vassall A. Superspreading, overdispersion and their implications in the SARS-CoV-2 (COVID-19) pandemic: a systematic review and meta-analysis of the literature. BMC Public Health. 2023;23: 1003. doi:10.1186/s12889-023-15915-1

27. Cori A, Valleron AJ, Carrat F, Scalia Tomba G, Thomas G, Boëlle PY. Estimating influenza latency and infectious period durations using viral excretion data. Epidemics. 2012;4: 132–138. doi:10.1016/j.epidem.2012.06.001

Figures

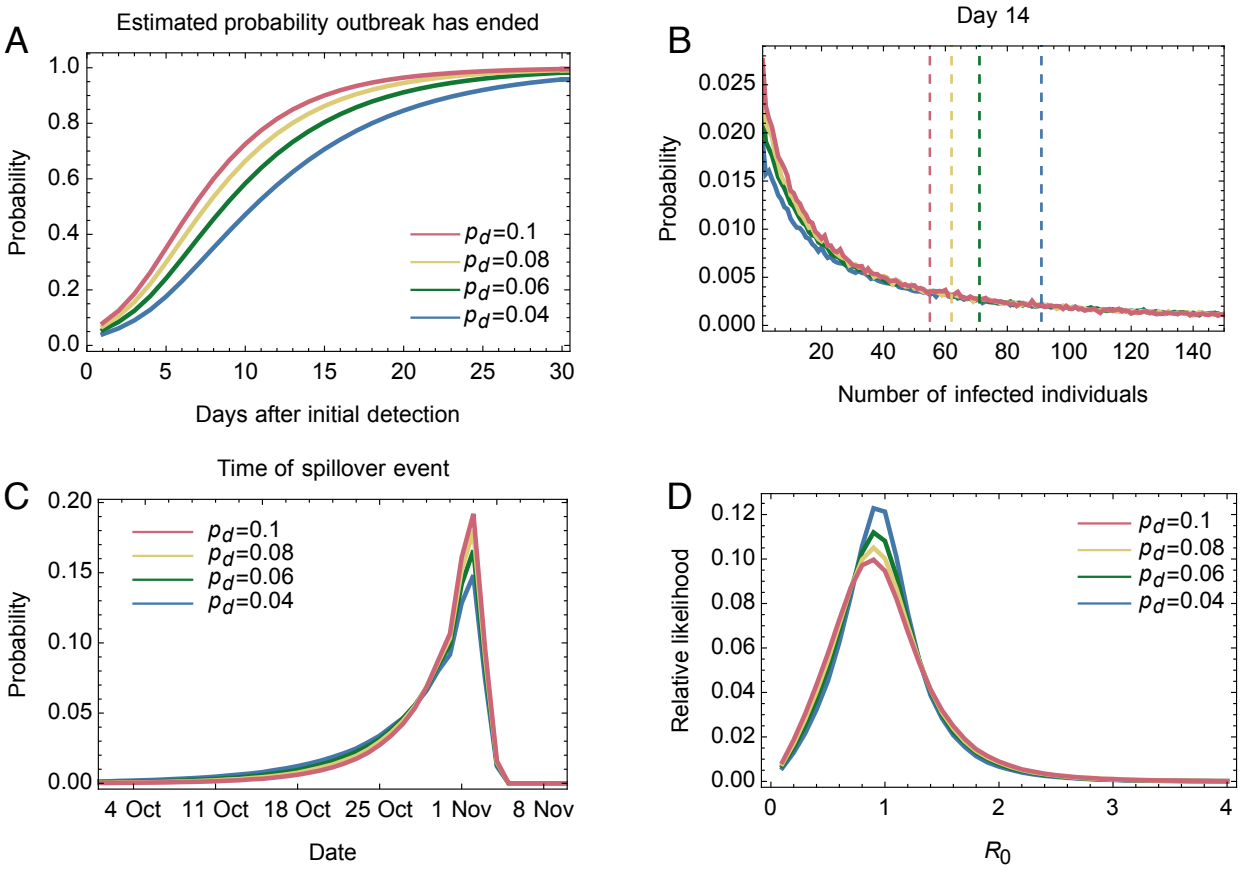

Figure 1: Statistics describing the influenza A/H1N2 spillover event calculated using our bootstrapping method. A. Time-dependent probability that the outbreak had died out. The value p_d describes the probability of a case of A/H1N2 influenza being detected. **B.** Calculated distribution describing the time at which the first case in the outbreak was infected. The detection of the first case was on 23rd November. **C.** Calculated distribution of the number of infected individuals 14 days after the date of the first detection, conditional on the outbreak having not died out. The vertical dashed lines show the median values of each distribution. **D.** Estimate of the epidemiological parameter R_0 for the influenza A/H1N2 virus involved in this spillover event.