
Estimating time to onset of swine influenza symptoms after initial novel A(H1N1v) viral infection

B. D. M. TOM¹, A. J. VAN HOEK², R. PEBODY², J. McMENAMIN³,
C. ROBERTSON³, M. CATCHPOLE² AND D. DE ANGELIS^{1,2*}

¹ *MRC Biostatistics Unit, Cambridge, UK*

² *Health Protection Agency, Centre for Infections, London, UK*

³ *Health Protection Scotland, Glasgow, Scotland, UK*

(Accepted 13 October 2010; first published online 19 November 2010)

SUMMARY

Characterization of the incubation time from infection to onset is important for understanding the natural history of infectious diseases. Attempts to estimate the incubation time distribution for novel A(H1N1v) have been, up to now, based on limited data or peculiar samples. We characterized this distribution for a generic group of symptomatic cases using laboratory-confirmed swine influenza case-information. Estimates of the incubation distribution for the pandemic influenza were derived through parametric time-to-event analyses of data on onset of symptoms and exposure dates, accounting for interval censoring. We estimated a mean of about 1·6–1·7 days with a standard deviation of 2 days for the incubation time distribution in those who became symptomatic after infection with the A(H1N1v) virus strain. Separate analyses for the <15 years and ≥15 years age groups showed a significant ($P < 0\cdot02$) difference with a longer mean incubation time in the older age group.

Key words: Influenza, pandemic.

INTRODUCTION

Infection with the novel influenza A(H1N1v) virus emerged in Mexico in March 2009 rapidly spreading worldwide. The WHO reports that as of 17 October 2009 there have been over 414 000 laboratory-confirmed cases of pandemic influenza and almost 5000 deaths worldwide (http://www.who.int/csr/don/2009_10_23/en/index.html). However, to date, relatively little is known about the natural history of the infection with A(H1N1v) virus. Knowledge of its incubation period, the time from infection to onset of

symptoms, is crucial to the understanding of the mechanism of transmission; to the recommendation of control measures such as contact tracing and quarantine; and to the formulation of transmission models aimed at predicting the healthcare burden [1–5].

At the beginning of the current pandemic, the incubation period had been generically quoted to be either between 1 and 4 days (CDC, 5 May 2009) or between 1 and 7 days [6]. A rigorous systematic review of the incubation period of acute respiratory viral infections then led to the adoption of a median incubation time of 1·4 days for influenza A [2] in some of the work on A(H1N1v) transmission modelling (e.g. [8]). A few attempts were also made to estimate the incubation period of the pandemic H1N1 2009

* Author for correspondence: Dr D. De Angelis, MRC Biostatistics Unit, Robinson Way, Cambridge CB2 0SR, UK.
(Email: daniela.deangelis@mrc-bsu.cam.ac.uk)

strain directly, by using information on exposure and symptom onset times from outbreaks [9–12]. These attempts reported incubation times of 24 h, a mean incubation time of 2 days and a median of around 3 and 2 days, respectively. These results are, however, mainly based on limited data with no attempt at explicitly modelling them. More considered is the work by Lessler *et al.* [13], who estimated the distribution of the incubation time from a school outbreak. The sample size is, in this case, reasonably large ($n=119$), but the sample is mainly composed of young adults in the 14–19 years age range. Similarly, Tuite *et al.* [14] formally address the problem, suggesting a mean incubation time of 4.3 days; much longer than previously estimated. A thorough characterization of the distribution of A(H1N1v) incubation period is, therefore, still warranted.

In the current study, information on time of exposure to the A(H1N1v) virus and on onset of symptoms from laboratory-confirmed cases collected from the Health Protection Agency (HPA) in England and Health Protection Scotland in the period May–July 2009 is used to more fully characterize the incubation period distribution.

METHODS

Data sources

From the HPA's national influenza database used to monitor cases during the containment phase of response (FluZone), we extracted all laboratory-confirmed novel A(H1N1v) cases in England with known date of onset of symptoms and which additionally had some degree of information available on the timing of exposure to the novel A(H1N1v) virus (either date of first or last exposure or ideally both these dates). The data obtained from FluZone were checked for completeness and consistency, and clarification on apparent anomalies were sought from the HPA units that had recorded the data. Data obtained from FluZone covered the period from 22 May 2009 up to 10 July 2009, after which FluZone was no longer used for monitoring the swine influenza pandemic. Furthermore, data on other confirmed cases with onset and exposure information available, but not in FluZone, were obtained from additional sources; specifically Regional HPA units in England and from Health Protection Scotland over the same time period. These other sources included data on investigated clusters linked with schools, a social gathering

and a hospital. They provided an additional 63 confirmed swine influenza cases.

Statistical analyses

Exact times (in days) to symptom onset from infection by the novel influenza A(H1N1v) virus were only known for confirmed cases who had point (i.e. single day) exposure contact with known index cases and those whose first known exposure date coincided with their date of onset (i.e. those with an exact exposure date). For the remaining cases (who were the majority) less precise data were available on the incubation time, which could only be assessed to fall within an interval (interval-censored incubation times) or to be longer (right-censored incubation times) or shorter (left-censored incubation times) than a specific length of time. We employed parametric survival analysis to analyse and characterize the data on incubation times. Models based on assuming log-normal, Weibull and gamma distributions for the incubation times were fitted and, from their estimated distributional parameters, means, medians and standard deviations for the incubation times were calculated for symptomatic swine influenza cases. A non-parametric maximum-likelihood estimator (NPMLE) of the incubation distribution was also constructed using the R package MLEcens [15] assuming that the boundaries of the interval-censored incubation times are opened on the left and closed on the right. All analyses were performed using the R statistical package [16].

Cases for which apparent inconsistencies in the data could not be resolved were excluded from analyses. Analyses were performed on various subsets of the collated data, beginning with the subset of cases for which exact incubation times could be obtained through knowledge of the exact exposure dates. This was then followed by the addition to this initial subset of those cases with both distinct first and last dates of exposure available. For these the exposure was only known to have occurred within an interval, thus producing interval-censored incubation times. Those with only a first date of exposure available (left-censored incubation times) were then added to the enlarged subset; and finally those with only a last exposure date known (right-censored incubation times) were included to constitute the full use of all available data (see Table 1). In addition, analyses based on FluZone and non-FluZone data were performed; as were age-stratified analyses in which age was dichotomized into

Table 2. Background characteristics of patients in the FluZone database and additional data sources

Data source	FluZone: no exposure information (<i>n</i> = 7280)	FluZone: some exposure information (<i>n</i> = 323)	Additional sources: some exposure information (<i>n</i> = 63)	<i>P</i> value of test for differences between sources
Sex				
Female	3370 (48.44 %)	162 (50.15 %)	32 (50.8 %)	χ^2 test: 0.782*
Male	3587 (51.56 %)	161 (49.85 %)	31 (49.2 %)	
Missing	323	0	0	
Age, yr				
Mean (s.d.)	19.08 (13.86)	20.69 (14.68)	22.71 (12.91)	ANOVA: 0.016*
Number	7227	322	63	
Missing	53	1	0	

* Ignoring missing data when performing the tests for difference.

NPMLE was between 2.043 days and 2.137 days. The estimated median incubation times to onset of symptoms (i.e. the times by which 50% of the confirmed cases are expected to show symptoms) for the entire dataset under the Weibull and gamma models were 0.963 (95% CI 0.765–1.174) and 0.948 (95% CI 0.717–1.158) days, respectively. The corresponding 95th percentiles were 5.653 (95% CI 4.759–6.832) and 5.646 (95% CI 4.889–7.003) days, respectively.

Analysis performed using only non-FluZone data gave a mean incubation time of around 2 days (s.d. = 1 day). Results from the age-stratified analyses are also shown in Table 1. From these results it appears that there is a difference between the mean incubation times for individuals aged <15 years and those aged ≥ 15 years, irrespective of the model used. Mean differences from the Weibull and the gamma models were 0.688 (95% CI 0.129–1.246) and 0.792 (95% CI 0.284–1.300) days, respectively. These differences are significant at the 0.02 level. Additionally, the estimated median incubation times from these two models were 0.365 (95% CI 0.124–0.659) and 0.248 (95% CI 0.012–0.521) for those aged <15 years; and 1.308 (95% CI 1.041–1.569) and 1.350 (95% CI 1.032–1.603) for those aged ≥ 15 years.

Figure 1 plots the non-parametric and parametric cumulative distribution functions obtained from analysing the full dataset of 335 confirmed cases, showing the cumulative proportion of cases developing symptoms by each day. These estimates are similar and suggest that the parametric Weibull and gamma models fit the data reasonably well. Figure 2 shows the bar plots of the probabilities of symptom onset each day after infection, up to day 10, derived from the parametric Weibull and gamma incubation models.

Here it is difficult to differentiate between the models which is reflected in their similar estimates of the mean incubation time and standard deviation (Table 1). However, on fitting a three-parameter generalized gamma model [17] for which both the Weibull and gamma (as is the log-normal) models are special cases, there is an indication that the Weibull is marginally more appropriate, in terms of having similar parameter estimates to the generalized gamma model and a slightly larger log-likelihood than the gamma model.

DISCUSSION

There is still substantial amount of uncertainty regarding the incubation period for the 2009 pandemic A(H1N1v) virus. Earlier attempts to estimate the incubation period were primarily based on limited data or on peculiar samples. In this work, we have characterized the incubation distribution using data from 335 laboratory-confirmed cases, both adults and children, derived from HPA's FluZone database for England and from additional sources, including Health Protection Scotland. The size and composition of this sample and the fact that all individuals included were confirmed as having had A(H1N1v) virus allowed us the unique opportunity to more fully characterize the incubation distribution of the population of symptomatic individuals in the UK, and provide a more precise estimate of the mean incubation time. Moreover, the timing of the sample collection, early in the epidemic, might have improved the likelihood of accurate exposure reporting as the population was mostly naive and cases more directly related to travel-acquired infections.

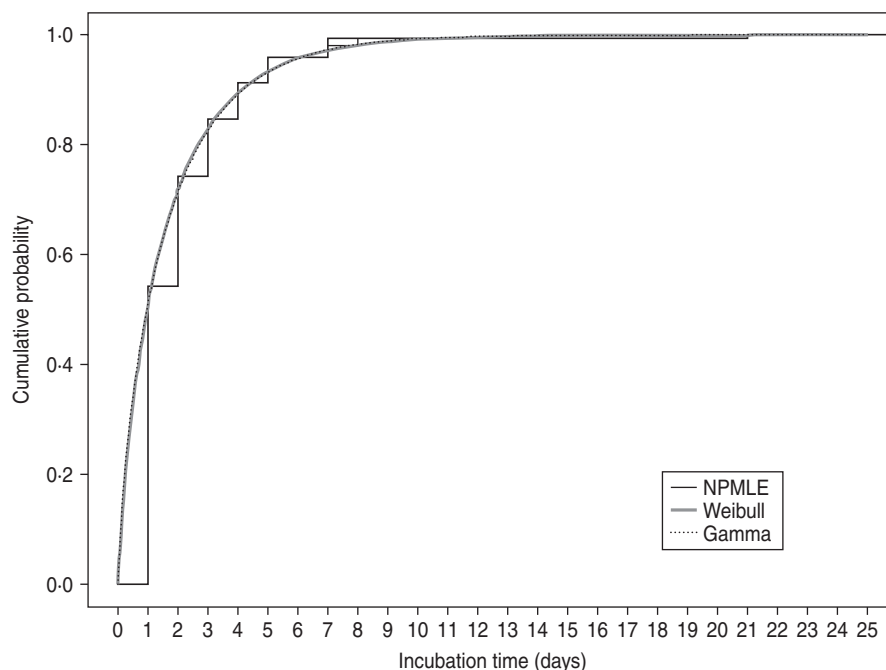


Fig. 1. Estimated cumulative distribution functions for the incubation times of the 335 confirmed symptomatic cases considered. NPMLE, Non-parametric maximum-likelihood estimator.

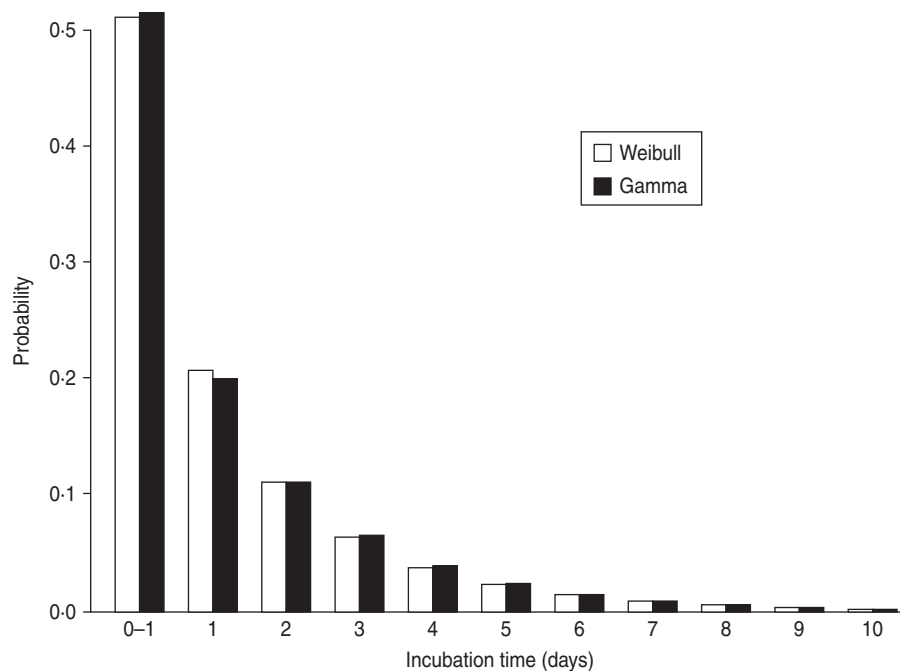


Fig. 2. Estimated probability of symptom onset each day after infection, up to day 10, for the 335 confirmed symptomatic cases considered.

We obtained a mean incubation time for symptomatic individuals of 1.6–1.7 days (s.d. = 2 days). The median incubation time was found to be between 0.9 and 1 day. Our estimated mean and median

incubation times are somewhat comparable to the median incubation time of 1.4 days obtained for influenza A [2], and our results for individuals aged ≥ 15 years agree with those obtained using data from

a New York City school outbreak on mainly young adults aged 14–19 years [13]. Additionally, our estimates are consistent with the earlier mentioned times of 24 h and 2 days for the UK pandemic, although lower than the suggested median of around 3 days reported for the pandemic in Spain [11]. On the other hand, they seem at odds with the mean incubation time of 4.3 days reported by Tuite *et al.* [14] using a log-normal distribution but without directly accounting for the interval-censored nature of the data.

Our study may suffer from selection bias as those positive cases who were not laboratory-confirmed may be different from those positive cases whose swabs were sent for laboratory confirmation; e.g. in the severity of their illness, with those laboratory-confirmed having a more severe illness. However, there was, in general, good ascertainment and laboratory testing of suspected cases up to the 10 July 2009, after which the FluZone database was no longer used for monitoring the swine influenza pandemic. As most of our cases were from the period before 10 July 2009, we are reasonably satisfied that there were few positive cases that developed symptoms and were not tested. Another possible source of selection bias in our study was the use of laboratory-confirmed cases that provided exposure information only. This group of cases may systematically differ from those laboratory-confirmed cases that had no recorded exposure details in ways that may be associated with time to symptoms from infection. However, it is unclear how these cases without exposure information and without detailed contact information could have been used here. Moreover, no obvious difference in sex distribution was seen (Table 2) and only a very marginal statistical difference ($P=0.04$) of 1.6 years was noted between the mean ages.

In this study we have assumed that the date of onset of symptoms was known exactly, while the exact date of infection with A(H1N1v) virus was only known for a minority of cases. This uncertainty in infection date for the majority of cases hampers analyses, and statistical methods (such as parametric distributional models) that handle interval-censored data are required to characterize the incubation distribution. Other approaches for estimating the incubation period distribution that do not rely on the time of infection, e.g. from serial interval data [18], were considered but found unfeasible. Additionally, it is quite plausible that cases may have inaccurately reported the exact date when their symptoms first arose, although it is unclear in what direction this

misreporting would be. If, in addition to uncertainty in the exact infection date, the onset of symptoms date was known to fall within an interval, then methods for handling such doubly interval-censored data would be required [5].

Our study, with all the provisos expressed above, provides a fuller characterization of the incubation distribution for a generic group of symptomatic cases infected with the novel A(H1N1v) virus. Not only have our analyses provided an estimate of the mean (or median) incubation time, which appears to be reasonable, but they have provided us with a well fitting characterization of the entire incubation distribution, based either on the Weibull or gamma parametric models, which also appropriately accounts for interval-censoring. A parametric representation of the incubation distribution will allow us to examine different aspects of the distribution (e.g. the 95th percentile) that are important for deciding containment strategies, identifying the source of infection and for the formulation of transmission models aimed at predicting the healthcare burden, as advocated, for instance, by Lessler *et al.* [2]. Furthermore, from our analyses we have found evidence for a difference in the mean incubation times for the <15 years and ≥ 15 years age groups, with a longer mean incubation time in the older age group. This finding, which has not been shown previously for influenza, may have implications on how these two age groups are treated in terms of the prioritizing of resources.

In conclusion, our characterization of the incubation time for the novel A(H1N1v) virus will hopefully contribute to a better understanding of the spread of this pandemic and aid in its control and management.

ACKNOWLEDGEMENTS

The authors thank the epidemiologists from the various Health Protection Agency units in England for their assistance in answering queries relating to the data. Brian Tom was financially supported by the UK Medical Research Council (MRC grant U.1052.00.009). Daniela De Angelis was funded by the UK Health Protection Agency and the UK Medical Research Council (MRC grant U.1052.00.007).

DECLARATION OF INTEREST

None.

REFERENCES

1. **Fraser C, et al.** Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* 2009; **324**: 1557–1561.
2. **Lessler J, et al.** Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infectious Diseases* 2009; **9**: 291–300.
3. **Melzl H, et al.** First sequence-confirmed case of infection with the new influenza A(H1N1) strain in Germany. *Eurosurveillance* 2009; **14**(18): pii = 19203.
4. **Odaira F, et al.** Assessment of secondary attack rate and effectiveness of antiviral prophylaxis among household contacts in an influenza A(H1N1)v outbreak in Kobe, Japan, May–June 2009. *Eurosurveillance* 2009; **14**(35): pii = 19320.
5. **Reich NG, et al.** Estimating incubation period distribution with coarse data. *Statistics in Medicine* 2009; **28**: 2769–2784.
6. **Centers for Disease Control and Prevention.** Interim guidance for clinicians on identifying and caring for patients with swine-origin influenza A (H1N1) virus infection (<http://www.cdc.gov/>). Accessed 5 May 2009.
7. **Sullivan SJ, et al.** 2009 H1N1 influenza. *Mayo Clinic Proceedings* 2010; **85**: 64–76.
8. **Lipsitch M, et al.** Use of cumulative incidence of novel influenza A/H1N1 in foreign travelers to estimate lower bounds on cumulative incidence in Mexico. *PLoS One* 2009; **4**: e6895.
9. **Kar-Purkayastha I, et al.** The importance of school and social activities in the transmission of influenza A(H1N1)v: England, April–June 2009. *Eurosurveillance* 2009; **14**(33): pii = 19311.
10. **Ghani AC, et al.** The early transmission dynamics of H1N1pdm influenza in the United Kingdom. *PLoS Currents: Influenza* 2009; November 16: RRN1130.
11. **Surveillance Group for New Influenza A(H1N1) Virus Investigation and Control in Spain.** New influenza A(H1N1) virus infections in Spain, April–May 2009. *Eurosurveillance* 2009; **14**(19): pii = 19209.
12. **Cao B, et al.** Clinical features of the initial cases of 2009 pandemic influenza (H1N1) virus infection in China. *New England Journal of Medicine* 2009; **361**: 2507–2517.
13. **Lessler J, et al.** Outbreak of 2009 pandemic influenza A (H1N1) at a New York City school. *New England Journal of Medicine* 2009; **361**: 2628–2636.
14. **Tuite A, et al.** Estimated epidemiological parameters and morbidity associated with pandemic H1N1 influenza. *Canadian Medical Association Journal* 2010; **182**: 131–136.
15. **Maathuis M.** MLEcens: computation of the MLE for bivariate (interval) censored data. R package 2007; version 0.1–2 (<http://www.stat.washington.edu/marloe/>).
16. **R Development Core Team.** R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (<http://www.R-project.org>).
17. **Farewell VT, Prentice RL.** A study of distributional shape in life testing. *Technometrics* 1977; **19**: 69–75.
18. **Kuk AYC, Ma S.** The estimation of SARS incubation distribution from serial interval data using a convolution likelihood. *Statistics in Medicine* 2005; **24**: 2525–2537.