




# The impact of climate and antigenic evolution on seasonal influenza virus epidemics in Australia

Edward K. S. Lam <sup>1</sup>, Dylan H. Morris <sup>2</sup>, Aeron C. Hurt<sup>3,4</sup>, Ian G. Barr<sup>3,4,5</sup> & Colin A. Russell <sup>6</sup>✉

Although seasonal influenza viruses circulate globally, prevention and treatment occur at the level of regions, cities, and communities. At these scales, the timing, duration and magnitude of epidemics vary substantially, but the underlying causes of this variation are poorly understood. Here, based on analyses of a 15-year city-level dataset of 18,250 laboratory-confirmed and antigenically-characterised influenza virus infections from Australia, we investigate the effects of previously hypothesised environmental and virological drivers of influenza epidemics. We find that anomalous fluctuations in temperature and humidity do not predict local epidemic onset timings. We also find that virus antigenic change has no consistent effect on epidemic size. In contrast, epidemic onset time and heterosubtypic competition have substantial effects on epidemic size and composition. Our findings suggest that the relationship between influenza population immunity and epidemiology is more complex than previously supposed and that the strong influence of short-term processes may hinder long-term epidemiological forecasts.

<sup>1</sup>Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. <sup>2</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. <sup>3</sup>WHO Collaborating Centre for Reference and Research on Influenza, VIDRL, Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia. <sup>4</sup>Department of Microbiology and Immunology, University of Melbourne, Parkville, VIC, Australia. <sup>5</sup>School of Applied Biomedical Sciences, Federation University, Churchill, VIC, Australia. <sup>6</sup>Department of Medical Microbiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. ✉email: [c.a.russell@amsterdamumc.nl](mailto:c.a.russell@amsterdamumc.nl)

Seasonal influenza virus epidemics are a substantial source of disease burden and result in ~650,000 deaths each year<sup>1</sup>. Four co-circulating subtypes/lineages of influenza viruses currently cause disease in humans: A/H3N2 (A/H3), A/H1N1 (currently A/H1pdm09, previously A/H1seasonal (A/H1sea)), B/Victoria/2/87-like (B/Vic) and B/Yamagata/16/88-like (B/Yam) viruses. The timing, duration and size of local influenza virus epidemics can vary substantially from year to year<sup>2,3</sup>, but the underlying causes of this variation are poorly understood. Better understanding of the factors that govern epidemic onset and magnitude could allow for accurate and timely epidemiological forecasts<sup>4</sup> and more efficient allocation of public health resources<sup>5</sup>.

In temperate regions of the Northern and Southern Hemispheres, influenza virus activity is most common in winter months, but the mechanistic basis of this seasonality remains unclear. Experimental studies demonstrated that reductions in temperature and absolute humidity enhance viral stability and aerosol transmission<sup>6–8</sup>. However, epidemics in tropical and subtropical regions often occur during periods of high temperature and humidity<sup>9</sup>.

Climatic fluctuations have been implicated as triggers for influenza epidemics in temperate regions. A study of state-level epidemiological data from the United States found that influenza epidemics sometimes follow 2-week periods of anomalously low absolute humidity<sup>10</sup>. Subsequent studies of epidemiological activity have found similar results using prefecture-level data from Japan<sup>11</sup>, city-level data from the New York Metropolitan Area<sup>12</sup> and region-level data from France<sup>13</sup>.

Influenza virus evolutionary dynamics are another theorised driver of influenza virus epidemiology. Within each type and subtype of seasonal influenza virus, new major antigenic variants arise every 3–8 years<sup>14,15</sup>. New variants partially escape the immunity induced by prior infections and vaccinations, rendering a higher fraction of individuals susceptible to infection. Epidemiological theory predicts that epidemics caused by a new antigenic variant should therefore be larger than epidemics of previously circulated variants<sup>16,17</sup>.

Antigenic change could also produce earlier and more spatio-temporally synchronous epidemics. When more individuals are susceptible, fewer transmission chains go stochastically extinct, so each new introduction of a virus into a population has a higher chance of causing an epidemic. Consistent with this, studies have suggested that antigenic change is associated with earlier epidemics in Israel<sup>18</sup>, and with more synchronous epidemics among cities in the United States<sup>19–21</sup>, Japan<sup>22</sup> and Australia<sup>23</sup>.

Studies of environmental and virological drivers of influenza virus epidemiology, including the studies referenced above, have been limited by three factors: (1) the reliance on influenza-like illness (ILI) data, (2) the aggregation of ILI or virologically confirmed data over large geographical scales (state/province/country) and (3) where virologically confirmed data are available, the use of data without subtype and antigenic variant-level resolution.

ILI data frequently include a wide variety of respiratory infections<sup>24</sup>, and limited laboratory characterisation obscures influenza virus type/subtype- and antigenic variant-specific patterns. These patterns become superimposed upon each other due to aggregation of ILI or virologically confirmed data to ecological scales (state/province/country) that sum over multiple local epidemics (county/city/town), which can individually vary substantially in timing, magnitude and influenza virus composition. Altogether, these sources of obfuscation make it difficult to disentangle local-level, antigenic variant-specific patterns, and critically investigate the impact of virus antigenic change.

Here, we use a 15-year data set of 18,250 typed, subtyped and antigenically characterised seasonal influenza viruses from the

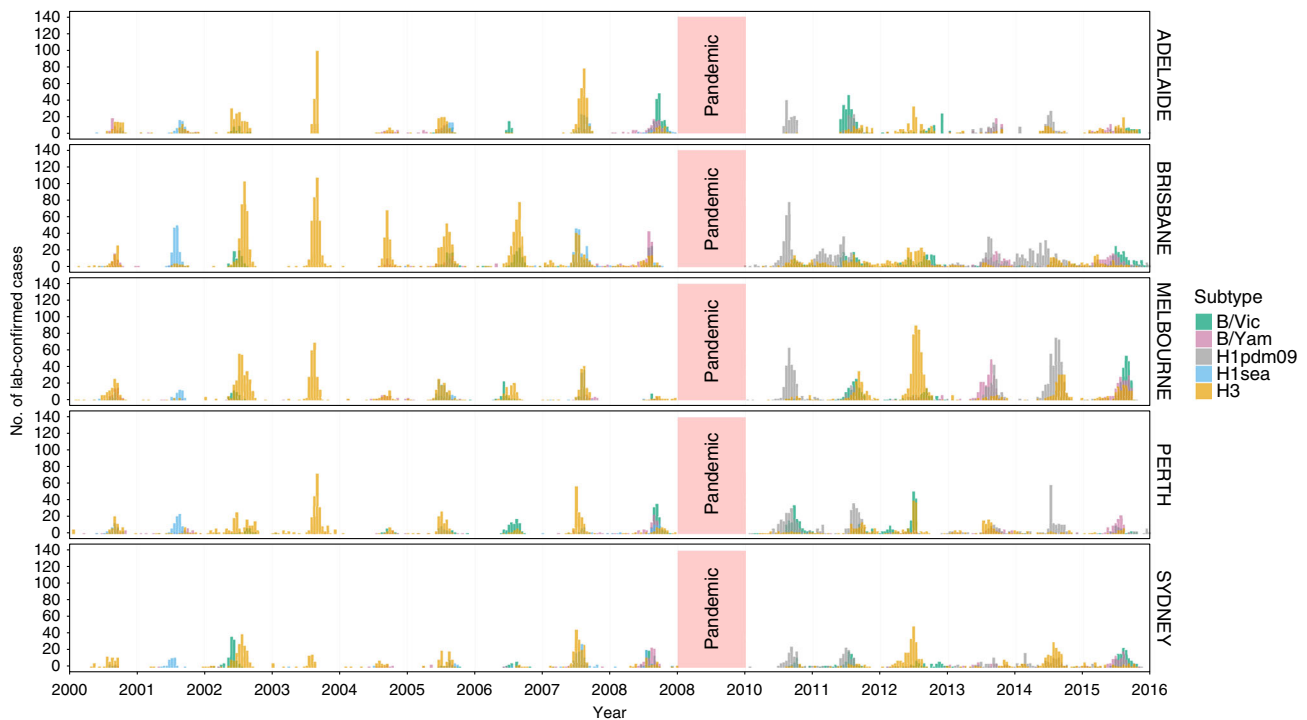
five most populous cities in Australia to investigate the impact of environmental and virological factors on the timing and magnitude of city-level influenza virus epidemics. We find that climatic fluctuations and virus antigenic change have no consistent effects on epidemic onset timing or size, while epidemic onset timing itself and heterosubtypic competition have substantial impacts on epidemic size and virus subtype composition. The lack of consistent effect of easily measured climatic and virus antigenic properties, and seeming dominance of noisy short-term transmission processes likely diminishes the feasibility of meaningful long-term influenza epidemic forecasting at local scales.

## Results

**Australia laboratory-confirmed influenza.** We aggregated 18,250 laboratory-confirmed and antigenically characterised cases of seasonal influenza viruses from 2000 to 2015 by 2-week (14-day) periods, creating a set of subtype- and antigenic variant-specific time series for the five most populous cities in Australia: Sydney (~5.5 million people), Melbourne (~5.0 million), Brisbane (~2.4 million people), Perth (~2.3 million) and Adelaide (~1.4 million) (Fig. 1). We excluded all virus cases from the 2009 season from all analyses because the 2009 A/H1N1 virus pandemic was atypical compared with seasonal epidemics and likely to be driven by different processes, affecting both epidemic dynamics and data collection of A/H1pdm09, as well as the other subtypes. Using a Poisson count detection method (see ‘Methods’), we identified periods of sustained, above-baseline levels of epidemic activity for each antigenic variant in each city. To facilitate comparisons among cities, we calculated the laboratory-confirmed incidence per 10<sup>6</sup> individuals using the annual estimated resident population values of each city<sup>25</sup>.

Epidemic magnitude and the most common virus subtype varied substantially among cities (Fig. 1). For example, during the 2002 season, A/H3 and B/Vic viruses were the most common strains in both Brisbane and Sydney. Absolute A/H3 virus incidence in Brisbane was much higher than in Sydney (186 vs 38.0 cases per 10<sup>6</sup> individuals), as was absolute B/Vic incidence (40.3 vs 22.7 cases per 10<sup>6</sup> individuals). But B/Vic had a substantially higher relative incidence in Sydney than in Brisbane (37% of all cases, vs only 18%). In some seasons, a virus antigenic variant caused a major epidemic in one or more cities, but failed to produce any observable above-baseline activity in another city. For example, in 2006, the A/Wisconsin/67/2005 (H3N2) virus variant caused epidemics in Brisbane, Perth and Melbourne, while above-baseline levels of activity were completely absent in Adelaide.

**Effect of climatic factors.** Epidemic onset timing varied substantially within and among cities and virus subtypes (Supplementary Fig. 1). Previously, Shaman et al.<sup>10</sup> showed that the 2-week period preceding the onset of state-level ILI epidemics in the United States was often marked by unusually low temperatures (T) or absolute humidities (AH). Fluctuations in these climatic factors from the historic averages expected for that specific day of the year (T' and AH', respectively) were anomalously large and negative when compared against a bootstrapped distribution of random samples from the historical records of observed daily climatic fluctuations recorded over wintertime (defined as 1 October–28 February). Following the same bootstrap-sampling method (see ‘Methods’) and aggregating epidemics across all five Australian cities, there were no statistically significant differences (all  $P > 0.05$ , see Supplementary Table 1) between the bootstrapped distribution of random samples of typical wintertime fluctuations (1 April–31 August for Australia) and the observed fluctuations in anomalous temperature and absolute humidity



**Fig. 1** Number of laboratory-confirmed seasonal influenza virus infections from 2000 to 2015 for the five largest cities in Australia. Cases are aggregated by 2-week periods, stratified by city and coloured by subtype/lineage.

over the 2-, 4- and 6-week periods immediately prior to the onset of the earliest epidemics from 2000 to 2015 (excluding 2009, 15 years  $\times$  5 cities = 75 epidemics in total) (Fig. 2). Individual city-by-city analyses (Supplementary Fig. 2 and Supplementary Table 2) showed that there was substantial local variation but no consistent patterns. Epidemic onset times coincided with both high and low temperature and absolute humidity periods, and there were no statistically significant patterns in four of the five cities.

Even if anomalous fluctuations in temperature and humidity do not necessarily affect epidemic onset, climatic factors could have an impact on virus transmission<sup>7</sup> and overall epidemic size: for example, influenza mortality in New York Metropolitan Area was shown to be negatively associated with temperature and humidity<sup>12</sup>. Overall, epidemic incidence should depend strongly on the initial exponential growth phase of the epidemic, where transmission may be facilitated by favourable climatic conditions. We therefore investigated the impact of mean temperature and mean absolute humidity during each epidemic, as well as just the period from epidemic onset to the peak, on that epidemic's size. For both time periods considered, epidemic incidence was not associated with mean absolute humidity (Supplementary Fig. 3). We found that epidemic incidence was weakly negatively associated with the mean temperature during the epidemic and the period from start to the peak, but this relationship appears to be primarily driven by two instances, where small epidemics occurred during the early and warmer part of the season; on balance, the highly variable epidemic sizes observed over a range of climatic conditions, suggest that climatic factors have limited and noisy effects (Supplementary Fig. 3).

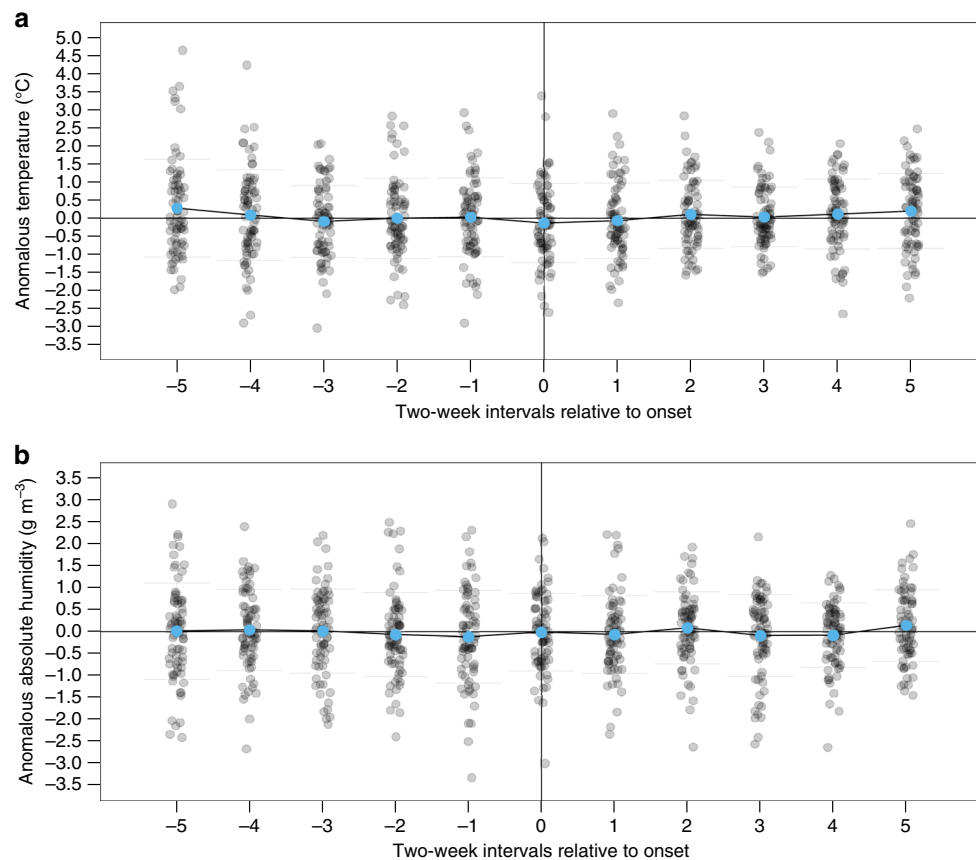
A recent study by Geoghegan et al.<sup>23</sup> estimated epidemic onset timings for influenza A virus epidemics in Australian postcodes for the seasons from 2007 to 2016. Despite the lack of subtype-level resolution, their data set is substantially larger (450,000 entries) than the one used here, and offers an opportunity to compare findings. We repeated our anomalous temperature and

absolute humidity analyses on the Geoghegan et al.<sup>23</sup> data set. As with our original data set, there were no consistent statistically significant relationships between climate anomalies and epidemic onset (Supplementary Discussion, Supplementary Tables 3, 4, Supplementary Figs. 4 and 5).

Other climatic factors have been proposed as drivers of influenza dynamics, notably relative humidity and rainfall<sup>6,9</sup>. We repeated the above analyses for relative humidity and rainfall. There were some city-level associations, but no consistent pattern and no pattern when aggregating across cities. Epidemic onset was not associated with statistically significant fluctuations in anomalous relative humidity and rainfall.

**Effect of antigenic change.** We next examined the effect of antigenic evolution on epidemic dynamics. Between 2000 and 2015, 7A/H3, 3A/H1sea, 1A/H1pdm09, 3 B/Vic and 5 B/Yam virus antigenic variants circulated in Australia. All A/H1pdm09 virus epidemics from 2009 to 2015 were excluded for this set of analyses for two reasons. First, we could not accurately estimate the size of the 2009 pandemic. Second, there was no subsequent, detectable antigenic change observed for A/H1pdm09 viruses during the study period. We normalised epidemic sizes (see 'Methods') to enable comparisons between cities. Stratifying by subtype/lineage, we compared the size of the first epidemic caused by an antigenic variant against the sizes of epidemics of the same antigenic variant in subsequent years (Wilcoxon two-sample test, Fig. 3). Contrary to the predictions of previous theoretical studies<sup>16,17</sup>, newly emerged antigenic variants caused epidemics, both larger and smaller than city-specific mean epidemic sizes, and there was no evidence of a consistent effect of antigenic change on epidemic size.

We also compared the timing of the first epidemic caused by an antigenic variant against the timings of subsequent epidemics (Supplementary Fig. 6) to test the hypothesis that new variants cause earlier epidemics. The range of onset timings was very



**Fig. 2 Climatic conditions around epidemic onset.** **a** Anomalous temperature  $T'$  and **b** absolute humidity  $AH'$  prior to and after epidemic onset across all five cities. Epidemic onset is marked by the vertical line at 0. For the earliest-onset epidemic in each season and city (15 years  $\times$  5 cities = 75 epidemics),  $T'$  and  $AH'$  for each time point are represented by grey points: a point below the horizontal line denotes that the value is lower than the 31-year city-specific mean. Blue points show the mean  $T'$  and  $AH'$  for that 2-week period for all epidemics across all cities within the study period. There were no periods with statistically significant reductions ( $P < 0.05$ ; Non-parametric bootstrapping) in  $T'$  or  $AH'$  from the 31-year averages.

broad, with epidemics starting from very early to late into the season, and there were no statistically significant differences in epidemic onset timing between new and extant variant epidemics.

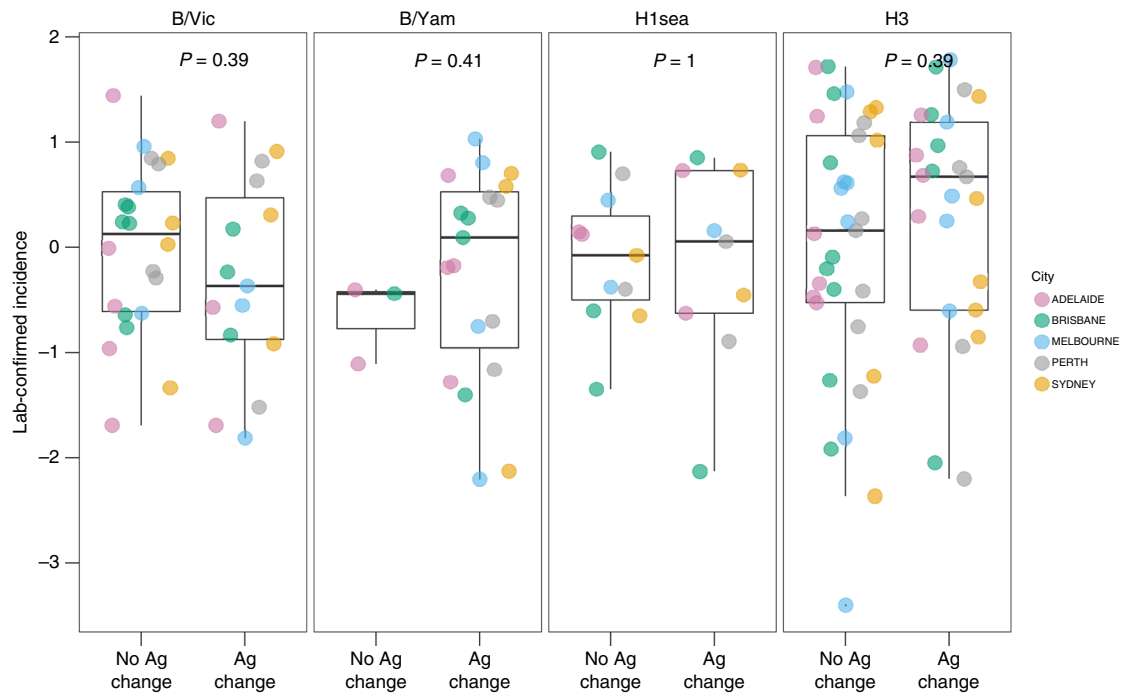
To investigate the impact of antigenic change on the spatiotemporal synchrony of epidemics, we examined the timing of epidemic activity across cities for years when a new major antigenic variant circulated in all five cities. New antigenic variants often failed to initiate epidemics across all five cities in a given year. We compared the synchrony of epidemics (defined as the reciprocal of the variance in epidemic onset timings) in the season in which an antigenic variant first emerges to the synchrony in subsequent seasons. There were no statistically significant differences in epidemic synchrony associated with antigenic novelty (Supplementary Fig. 7).

To check the robustness of this result, we repeated these analyses using estimated-onset timings from Geoghegan et al.<sup>23</sup>. There was again no discernible effect of antigenic change on the timing or synchrony of epidemics (Supplementary Discussion).

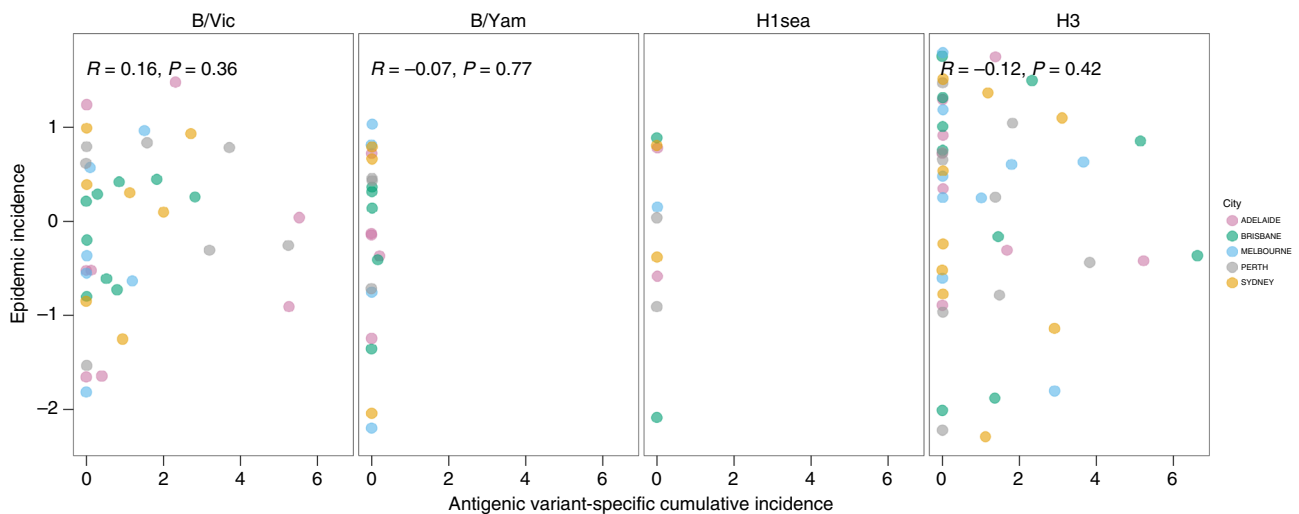
**Effect of prior immunity.** After an antigenic variant causes an epidemic in a city for the first time, the accumulated population immunity to that variant should lead to smaller subsequent epidemics, and eventually render further epidemics of that variant less likely. For each epidemic caused by a given antigenic variant, we investigated the relationship between that epidemic's size and the cumulative number of cases caused by that antigenic variant in preceding seasons. To account for differences in population

size and surveillance intensity among cities, we normalised epidemic and cumulative case counts by the city-specific mean epidemic size. Antigenic variants that emerged prior to the start of the study period, such as A/Moscow/10/99 (A/H3) and A/New Caledonia/20/99 (A/H1sea) and all A/H1pdm09 epidemics from 2009 to 2015 were excluded from this analysis, since it was not possible to calculate cumulative case counts for them. Specific B/Yam antigenic variants rarely caused more than one epidemic in a given city, but specific antigenic variants of A/H3 and B/Vic viruses caused repeated epidemics in the same city. For A/H3 and B/Vic viruses, epidemic size and cumulative prior incidence were not correlated (Pearson's correlation test, Fig. 4).

The accumulation of population immunity should also reduce the probability of successful epidemic initiation, making epidemics, regardless of size, less likely to start after an antigenic variant has already caused an epidemic in that city. For B/Vic and A/H1sea viruses, binary logistic regression showed non-significant associations between the cumulative incidence over prior seasons and the probability of successful epidemic initiation (all OR  $< 1$ ; all  $P > 0.05$ , Supplementary Fig. 8 and Supplementary Table 5). This partially resulted from the small number of A/H1sea epidemics during the study period, most of which were caused by newly emerged antigenic variants. However, B/Yam and H3 viruses showed significant negative relationships between cumulative prior incidence and epidemic probability, suggesting that prior incidence may have a substantial impact on the probability of successful epidemic initiation.



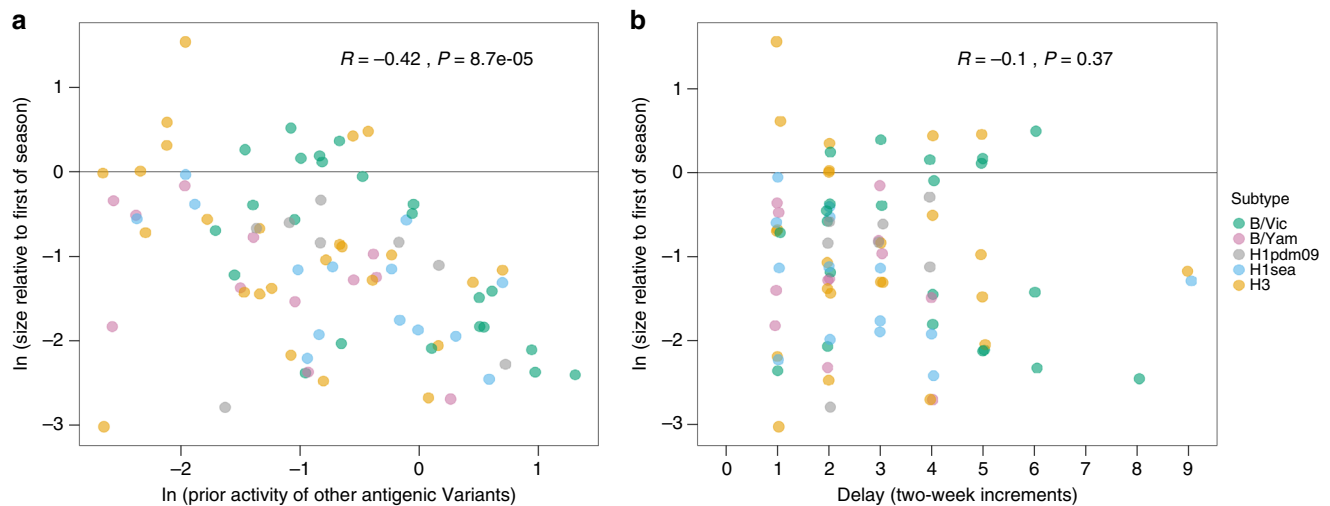
**Fig. 3 Effect of antigenic change on epidemic incidence.** Epidemic incidence was compared between seasons associated with and without the epidemic-level circulation of a new major antigenic variant. Within each subtype, the incidence for individual epidemics was log-transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. *P* values are from Wilcoxon two-sample tests ( $n = 37, 26, 22$  and  $63$  for B/Vic, B/Yam, A/H1sea and A/H3, respectively). Each point corresponds to one epidemic in a city, and the box plots show the median, first and third quartile of the transformed values, and range.



**Fig. 4 Effect of prior immunity on epidemic incidence.** Within each subtype, the incidence for individual epidemics was log-transformed and subtracted by the city-specific mean of log incidence, to allow for comparison between cities. Antigenic variant-specific cumulative incidence was measured relative to the city-specific mean epidemic size, where 1 is equivalent to the mean epidemic incidence. *r*- and *P* values are from Pearson’s correlation tests ( $n = 37, 20, 9$  and  $45$  for B/Vic, B/Yam, A/H1sea and A/H3, respectively). Antigenic variants of B/Yam rarely initiated multiple epidemics during the study period, and it was not possible to calculate a correlation coefficient for A/H1sea because the one new antigenic variant to emerge during the study period caused only a single epidemic per city.

**Aggregating across subtypes.** There may be subtype/lineage-specific differences in the effect of antigenic change and prior immunity. Notably, B/Yam antigenic variants typically cause only one epidemic per city. We repeated these analyses with epidemics aggregated together, across all subtypes and cities to increase statistical power (see the project Github repository for the analyses and code). As before, there were no statistically significant differences in the magnitude of epidemics between the

first and subsequent epidemics of an antigenic variant, or any association between epidemic size and the cumulative incidence over prior seasons. Binary logistic regression showed that the probability of successful epidemic initiation may be moderately reduced by the cumulative incidence over prior seasons. Our findings were robust to the method of normalisation used to allow for comparison between cities and subtypes/lineages (see ‘Methods’).



**Fig. 5 Effect of competition among subtypes on epidemic incidence.** The relationship between the size of an epidemic and **a** the amount of prior activity of all other antigenic variants and subtypes, and **b** the delay in epidemic onset. The size of each epidemic, relative to the earliest epidemic of that season, was log-transformed: the horizontal line at 0 denotes that the size of an epidemic is equal to that of the earliest epidemic of that season and city. In panel **(a)**, prior activity by other subtypes within the same season was measured relative to the city-specific mean epidemic size. In panel **(b)**, delay in epidemic onset was measured relative to the onset timing of the earliest epidemic of that season.  $r$ - and  $P$  values are from Pearson's correlation tests ( $n = 82$ ).

**Effect of competition among subtypes.** Competition among virus subtypes for hosts should create a first-mover advantage for the first subtype to sustain above-baseline epidemic activity in a city in a given season. Subsequent epidemics of other subtypes within that same season should therefore be reduced in size. We considered two proxies for this kind of intersubtypic interference: the cumulative amount of epidemic activity prior to the onset of a subtype's epidemic and the lag between the focal epidemic and the season's earliest epidemic. To allow for comparisons across cities and subtypes, we normalised log-epidemic case counts by subtracting off the city- and subtype-specific mean log-epidemic case count. There was a strongly negative and statistically significant correlation between prior epidemic activity and epidemic size (Pearson's correlation test,  $r = -0.420$ ;  $P = 8.7e-5$ , Fig. 5). An important caveat is that seasonality in the transmission rate could result in epidemics that start later in a season being smaller than those that started earlier, regardless of intersubtypic competition.

**Joint contributions of climatic and virological factors.** Whilst the magnitude of the effects of the climatological and virological factors may be individually subtle, it could be the case that they are only able to affect observable changes on the magnitude and timing of epidemics when acting in concert, or that large effects in opposing directions mask each other. We used a Bayesian multilevel regression model to identify which putative predictor variables affected epidemic incidence, and estimate posterior distributions for their effects on epidemic size. The model included the following variables: antigenic change, cumulative prior cases of the antigenic variant, mean absolute humidity during the epidemic, activity by other subtypes earlier in the season, epidemic start date and rainfall during the epidemic. Mean temperature during the epidemic was omitted as a predictor, since it was highly collinear with absolute humidity; analyses were subsequently repeated using mean temperature, and omitting absolute humidity with no substantial changes in the overall results.

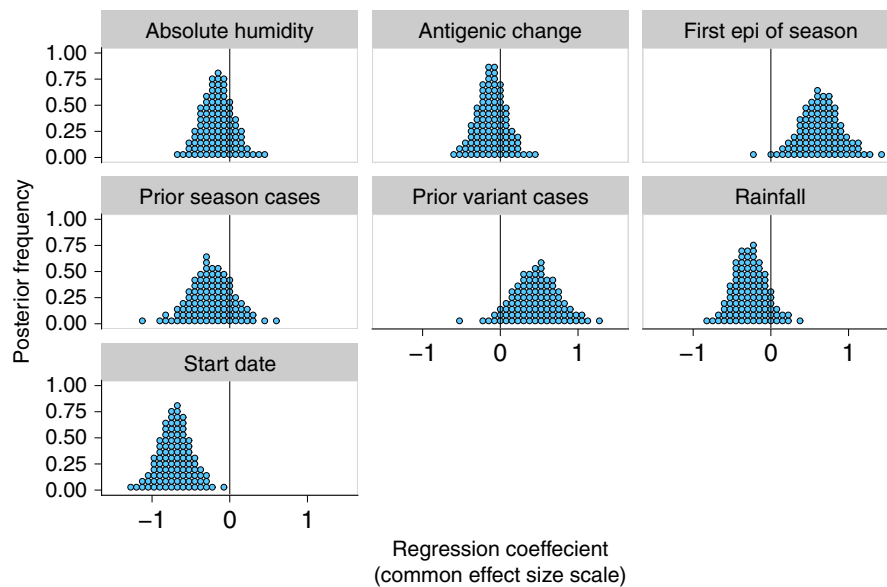
The model suggested that epidemics that were the first of the season or had early start dates should be modestly larger (Fig. 6). Start date had the largest estimated effect and the clearest

posterior support for a non-trivial effect size. Posterior modes for the mean effects of antigenic change and absolute humidity across subtypes were near zero (Fig. 6), with tight credible intervals (95% credible intervals:  $(-0.56, 0.27)$  for absolute humidity,  $(-0.50, 0.30)$  for antigenic change). Prior cases of an old variant given no antigenic change (95% credible intervals  $(-0.26, 0.85)$ ), prior cases of all variants for non-first epidemics (95% credible intervals  $(-0.84, 0.33)$ ) and rainfall during the epidemic (95% credible intervals  $(-0.68, 0.20)$ ) also showed no strongly discernible effects, though with less posterior certainty. The model could not explain much of the variation in the data: the median-estimated standard deviation of log epidemic size about the expected log size is 0.77 (95% credible intervals  $(0.67, 0.90)$ ). Since  $\exp(0.77)$  is  $\sim 2.15$ , this implies that it is not unusual to see epidemics half or twice the expected incidence. The model estimated that the effects were very similar across subtypes (Supplementary Fig. 14, median-estimated SDs for the distribution of subtype-specific effect sizes about the overall mean effect size near zero, Supplementary Fig. 15). Only the effect of whether an epidemic was the first of the season showed meaningful heterogeneity: the model estimated that it is somewhat weaker for B/Vic than for other subtypes (Supplementary Fig. 14).

## Discussion

Based on city-level analyses of a subtyped and antigenically characterised influenza virus data set covering the five largest cities in Australia, we find that climate and antigenic novelty have limited effects on epidemic sizes. The results presented here suggest that, at least in temperate areas, epidemics are governed by factors other than host immunity at local scales, where global fitness advantages for new antigenic variants may not be realised. Conversely, competition for hosts among influenza virus types and subtypes has strong effects on local dynamics. The first virus subtype to establish above-baseline epidemic activity in a city and season typically dominates.

A recent study of fine-scale influenza epidemiology in Australia<sup>23</sup> showed that there was substantial heterogeneity among Australian cities in the activity of influenza A and B viruses. Our subtyped and antigenically characterised data set allowed us to confirm that further heterogeneity exists at the level of antigenic



**Fig. 6 Joint contributions of climatic and virological factors on epidemic incidence.** The mean effects across all subtypes were estimated using the Bayesian multilevel model. Predictors were mean-centred and scaled, so effect sizes are shown on a common scale.

variants. In particular, specific antigenic variants often cause large epidemics in some cities while not causing detectable activity in others.

While prior studies found that the onset of epidemics in the United States and France was preceded by a 2-week period of anomalously low absolute humidity<sup>10,13</sup>, we found no evidence for climatic effects when aggregating across the five Australian cities. Anomalous fluctuations in temperature and absolute humidity were sometimes positive, sometimes negative, but on average approximately zero. Importantly, the overall effect size reported by Shaman et al.<sup>10</sup>, after aggregating across all 48 contiguous states of the United States, was very small (with mean AH' being  $\sim -0.25 \text{ kg kg}^{-1}$  or  $-0.21 \text{ g m}^{-3}$ , compared against  $0 \text{ g m}^{-3}$ , the mean of the null distribution of historic wintertime values). About 55–60% of epidemics were preceded by negative AH' values: a moderate increase upon the null hypothesis being a baseline of 50%.

Shaman et al.<sup>10</sup> also found regional differences in the associations between fluctuations in absolute humidity and epidemic onset. Strong associations were found in the Southeastern United States but not in Western states. In Australia, there does not appear to be an aggregate effect at the country level, and there were no consistent patterns at the level of individual cities (Supplementary Fig. 2 and Supplementary Table 2). The small effect sizes and lack of consistency in climatic patterns across regions and cities in the United States and Australia may reflect the fact that climatic factors alone are unlikely to account for the differences in the patterns of influenza seasonality between temperate and tropical regions<sup>26</sup>.

Seasonal epidemic waves in the United States appear to begin in the Southern states, which have warmer and more humid climates<sup>21,27</sup>, casting some doubt on the role of low humidity as a trigger for influenza epidemics. Rather than acting as specific triggers, it is plausible that climatic factors are acting on longer timescales than the anomalous fluctuations reported by Shaman et al.<sup>10</sup> to more generally enhance transmission and increase incidence<sup>28</sup>. However, in Australia, epidemic size does not appear to be strongly associated with the mean temperature or absolute humidity over the epidemic period.

Given the interest in influenza virus as a model system for phylodynamics of a pathogen that consists of multiple co-existing

antigenic variants<sup>29</sup>, there is interest in understanding how competition between these related variants, typified by cross-immunity, shapes epidemiological dynamics. Studies have hypothesised that antigenic change should result in larger<sup>16,30–32</sup> and earlier<sup>18</sup> local epidemics, which exhibit greater spatiotemporal synchrony at the national level<sup>19–23</sup>. The sequential replacement of old antigenic variants by new ones is indicative that antigenicity and population immunity are important for the global-level phylodynamics of influenza viruses. In contrast, at the local level, we find for A/H3 and B/Vic viruses that neither antigenic change nor the accumulation of antigenic variant-specific immunity are strong drivers of epidemic size, though accumulating variant-specific immunity may moderately reduce the probability of successful epidemic initiation.

It is striking that individual antigenic variants of A/H3 and B/Vic viruses are capable of re-invading the same city multiple times over consecutive years, despite a lack of substantial antigenic change. A/H1pdm09 viruses had previously been shown to cause repeat epidemics without antigenic change<sup>33,34</sup>, but our study establishes that this occurs for multiple types and subtypes of human influenza. One possible explanation for the lack of evidence for the year-on-year depletion of susceptible hosts is that influenza virus infection often fails to confer strain-specific and effective immunity. In some individuals, antigenic seniority and existing immunity against previously encountered antigenic variants may suppress novel strain-specific antibody responses, leading to only modest specific protection against reinfection<sup>35,36</sup>. Similarly, vaccine trials suggest that multiple exposures can be required in order for children to become seropositive sufficiently to protect themselves<sup>37</sup>. Potentially, multiple natural infections may also be needed to confer protective immunity<sup>38</sup>, particularly in children<sup>39</sup>, who effectively form a non-depleting pool of susceptibles.

There may also be substantial and previously unaccounted heterogeneity in individual susceptibility towards the same virus strain. The notion that population-level strain-specific immunity to influenza viruses is monolithic may be an artefact of the single-infection ferret models typically used to estimate antibody-mediated protection. In humans, there is substantial individual-to-individual variation in the antigenicity of amino acid escape mutations for influenza haemagglutinin<sup>40</sup>. Such heterogeneity

between individuals stems from their varied exposure histories to different influenza viruses. Unfortunately, age records for our data set were too incomplete to allow us to study age-specific heterogeneities in demographics, and attack rates between cities, and whether such patterns change over seasons.

Spatial and social connectivity structures among hosts in a city may also limit the spread of epidemics. Heterogeneous contact patterns between hosts can have a substantial impact on the resulting epidemiological dynamics<sup>41,42</sup>. Epidemics may be inherently frail processes: relatively minor human behavioural or environmental perturbations could prematurely terminate epidemics before they exhaust the pool of susceptible hosts, preserving a substantial number of susceptibles, and permitting subsequent epidemics of the same antigenic variant.

While our data set is substantially smaller (>450,000 vs 18,250 cases) than the one analysed by Geoghegan et al.<sup>23</sup>, and is thus more likely to be affected by noise in epidemic and surveillance processes, the differences between our findings and theirs highlight the importance of subtyping and antigenic characterisation, particularly for drawing conclusions about the effects of antigenic change. Geoghegan et al.<sup>23</sup> had cautiously suggested, given only virus-type data, that the 2009, 2012 and 2014 influenza A virus epidemics in Australia exhibited greater spatiotemporal synchrony potentially due to the emergence of the novel A/H1pdm09 subtype in 2009 and novel A/H3 antigenic variants in 2012 and 2014. However, with further subtype resolution and antigenic characterisation, we find that the majority of influenza A activity in Adelaide and Melbourne in 2014 was attributable to A/H1pdm09, rather than the (antigenically novel) A/H3; in fact, there was no above-baseline A/H3 activity in Perth. The fact that different virus subtypes caused these apparently synchronous epidemics implies that the epidemic synchrony described by Geoghegan et al.<sup>23</sup> was not due to the antigenic evolution or regional spread of a single virus strain.

Apart from competition between antigenic variants, previous epidemiological studies have hypothesised the existence of heterosubtypic competition where prior infection by a virus of one subtype is negatively associated with subsequent infection by a virus of another subtype<sup>43,44</sup>. In agreement with a previous US study of national-level ILI activity augmented with limited virus subtyping<sup>45</sup>, we also find evidence for a first-mover advantage and competition to infect hosts within a city, where the subtype or type that initiates above-baseline levels of activity first is most likely to have the largest epidemic of that season.

There are multiple caveats to our study that merit explicit consideration. The most important ones derive from our use of passive surveillance data that might not accurately reflect true underlying influenza virus activity. For example, surveillance intensity could plausibly vary between cities and years. While variation in surveillance efforts is evident among cities, there was no evidence of systematic increases or decreases in the number of laboratory-confirmed cases, or changes to surveillance practices within each city during the study period. Despite this, the longer duration of epidemics recorded after 2009 could be indicative of enhanced surveillance in the post-pandemic era: to mitigate this possibility, we repeated our analyses on the effect of antigenic change on epidemic size, splitting between pre- and post-pandemic eras and epidemic sizes normalising by their respective era-specific means. In either era, there was no consistent effect of antigenic change on epidemic size, with the caveat that splitting across eras reduced the number of observations in each era and thus our statistical power (Supplementary Fig. 18).

The intensity of surveillance could also vary over the course of an epidemic. For example, sentinel physicians could become more likely to submit samples for further testing as an epidemic unfolds, or conversely, testing could prematurely cease as facilities

become overwhelmed with samples. Despite being unable to definitively rule out the former scenario, the latter is unlikely to affect our data. If reporting ceased after a certain number of samples had been tested, the distribution of epidemic sizes would be truncated, and each epidemic would be unlikely to have an exponentially declining tail. No such patterns exist in our data.

The intensity of surveillance could also potentially vary across subtypes and lineages. The mean age of infection for A/H3 is greater than influenza B<sup>46</sup> viruses, and healthcare-seeking behaviour may differ between adults, parents with children and children. Furthermore, it is commonly thought that A/H3 virus infections result in more severe clinical presentations and greater risk of mortality<sup>47</sup> than influenza B viruses, potentially resulting in differences in the likelihood of detection by a sentinel health practitioner, though this may not be the case (see ref. 48).

Another important caveat is that while we were able to include antigenic data in this study, these data were all derived from haemagglutination inhibition (HI) assays. HI assays do not measure virus antigenic changes that occur away from the receptor-binding site, and thus likely represent an incomplete picture of antigenic change. Reference viruses and sera used in the haemagglutination inhibition assays can also impact the interpretation of the assay readout, and the HI data used in this study were therefore treated with caution (see 'Methods').

In this study, we attempted to identify associations between population susceptibility and epidemic incidence. Accurately quantifying the former is a complex challenge, so cumulative antigenic variant-specific epidemic incidence was used as a proxy, but that itself is subject to the limitations listed above. Besides natural infection, immunity can also be derived from vaccination, the contribution and effectiveness of which could not be determined due to a lack of temporally and geographically complete vaccination records over the study period. Regardless, we hypothesise that the impact of seasonal vaccination would be limited, particularly in the context of Australia, given the low uptake of vaccination<sup>49</sup>. Crucially, the uptake by children, who are important in driving local community transmission, is often below 10%<sup>50</sup>.

While our Bayesian multilevel model estimated negligible effects on epidemic size stemming from climatic factors and prior cases attributed to the same antigenic variant, the estimated credible intervals were not tight enough to rule out these effects conclusively (Fig. 6). However, our study suggests that climatic and antigenic factors are unlikely to be strong drivers of local influenza epidemiological dynamics. Indeed, the effects of these specific factors are dwarfed in magnitude by more generic epidemiological drivers: seasonality not directly captured by climate (measured by start date) and competition for hosts among subtypes (measured by whether an epidemic is the first of the season) (Fig. 6). We also find that even with all generic and specific factors considered, precise predictions of epidemic size remain difficult because of substantial noise in the local epidemic process.

Our Bayesian multilevel model for epidemic size avoids explicitly modelling underlying transmission processes, and may fail to fully capture the nature of the relationship (linear vs nonlinear) between transmission rates/ $R_0$  and the total cases in an epidemic. However, based on previous virus-transmissibility studies<sup>7</sup>, if climatic factors are strong drivers of epidemiological dynamics, we would expect the climatic variabilities observed in Australia to have a substantial impact on transmission rates, and produce detectable differences in epidemic size, but this is not the case.

Climatic drivers of seasonality and homosubtypic competition between virus antigenic variants are thought to be strong drivers of seasonal influenza epidemiology, but seasonal influenza virus epidemiological dynamics in major Australian cities appear to be



more substantially shaped by other factors, particularly the establishment of sustained virus-transmission activity, and subsequent competition among virus types and subtypes. This implies that the time horizon for meaningful forecasting of epidemic subtype composition is very short (days to weeks), and forecasting efforts aimed at longer-term predictions will require further insights into the dynamics of virus introduction and epidemic establishment, and into the accumulation of population immunity to seasonal influenza viruses.

**Methods**

**Australian surveillance data.** Influenza viruses from Australia were collected by the WHO Collaborating Centre (WHOCC) for Reference and Research on Influenza in Melbourne, Australia. The Melbourne WHOCC receives a subset of influenza-positive clinical samples collected by various sentinel surveillance systems across Australia throughout the year. The samples in this study were typed, subtyped and antigenically characterised by haemagglutination inhibition assay to the vaccine reference vaccine strain in use at the time of sample collection.

The data set consists of 18,250 influenza-positive cases, collected between 2000 and 2015 in the city of Brisbane, the city of Perth, the state of South Australia, the city of Sydney and the state of Victoria. The breakdown at the subtype/lineage level is as follows: A/H3 (7661), A/H1sea (1410), A/H1pdm09 (3987), B/Vic (3021) and B/Yam (2171). All of these correspond specifically to individual cities, except for the data from Victoria and South Australia. As of June 2015, 75 and 78% of the inhabitants of the states of Victoria and South Australia resided in the cities of Melbourne and Adelaide, respectively. We therefore treated the Victoria and South Australia data as representative of city-level patterns in those two major cities.

All epidemic activity of all subtypes for the 2009 season was excluded from all analyses because of the 2009 A/H1N1 pandemic. Unsurprisingly, patterns of virus circulation during the pandemic were anomalous compared with typical seasonal influenza virus epidemics, and potentially distortive of the patterns we sought to characterise.

**Estimation of epidemic timing.** The exact timing of interseasonal periods of sporadic activity and epidemic onset for each subtype is highly variable between years, even for individual cities, so it is necessary to determine the onset and end of each epidemic independently for each antigenic variant, season and city.

For each individual antigenic variant-specific time series, we used a Poisson count detection algorithm implemented in the Surveillance package in R<sup>51,52</sup> to distinguish periods of sustained epidemic activity from a background of sporadic interseasonal activity. We assume that the start of the calendar year falls sometime within the interseasonal period, which is justified by the scarce number of cases observed during this time of the year, and the fact that it is summertime in Australia. Making no further assumptions on the exact duration and timing for the interseasonal period or epidemic onset, starting at the beginning of the year, successive 2-week periods  $y_t$  are evaluated using the number of cases in each of the  $n$ -preceding 2-week periods  $\{y_{t-n}, y_{t-n+1}, \dots, y_{t-2}, y_{t-1}\}$  as reference values for sporadic activity. These reference values are used to predict a threshold value  $y_\alpha$ : if the observed number of cases  $y_t$  exceeds the threshold  $y_\alpha$ , the focal 2-week period is marked as the 2-week period of epidemic onset.

The Poisson count model assumes that the reference values  $y_t$  are identically and independently Poisson distributed with a mean of  $\lambda$  (Eq. (1)).  $\lambda$  itself has a Gamma distribution as a prior (Eq. (2)). From Eqs. (1) and (2), the posterior predictive distribution is a negative binomial distribution (Eq. (3)).

$$y_t \sim Po(\lambda) \tag{1}$$

$$\lambda \sim Ga(\alpha, \beta) \tag{2}$$

$$z|y_{t-n}, y_{t-n+1}, \dots, y_{t-2}, y_{t-1} \sim NegBin\left(\alpha + \sum_{i=1}^n y_i, \frac{\beta + n}{\beta + n + 1}\right) \tag{3}$$

The threshold value  $y_\alpha$  can then be calculated using quantile parameter  $\alpha$ , where  $y_\alpha$  is the smallest value that satisfies Eq. (4).

$$p(y \leq y_\alpha) \geq 1 - \alpha \tag{4}$$

We used the same algorithm to identify the end of an epidemic. Starting at the end of the year, successive 2-week periods, in the backward direction, are evaluated using the number of cases in each of the  $n$  following the 2-week period as reference values  $\{y_{t+1}, y_{t+2}, \dots, y_{t+n-1}, y_{t+n}\}$ .

During interseasonal periods, where there were often many 2-week periods reporting no cases, an isolated 2-week period with sporadic activity can be misconstrued as the onset of an epidemic. To reduce the impact of outliers in the time series and increase specificity of the detection algorithm, we first applied the 4253H, twice nonlinear data-smoothing algorithm<sup>53</sup>, which is a compound smoother consisting of multiple running medians.

We tested a variety of  $n$ - and  $\alpha$ -parameter values, and chose  $n = 3$  and  $\alpha = 0.12$  for the analyses presented in the text as a good compromise between sensitivity and

specificity in the identification of all of the epidemics within the time series and their individual onset and end timings, which were confirmed by visual inspection. The results of these analyses are also robust to alternative parameter values and corresponding changes to the sensitivity and specificity of the Poisson count detection algorithm (see Supplementary Discussion for sensitivity analyses).

Aggregation of cases by 2-week periods was deemed necessary, to smoothen the time series in light of the relatively low number of cases within the data set; this relatively long timescale could however potentially obscure fluctuations in weather that occur at shorter scales. Whilst weekly time series were appreciably noisier, we found a high degree of correspondence in the estimated epidemic onset and end timings with values calculated from data aggregated by 2-week periods: indeed, our results were robust to aggregation by week (see Supplementary Discussion for sensitivity analyses).

We deemed an antigenic variant to have failed to cause an epidemic if, within a season, the algorithm was unable to define an epidemic period; we confirmed all putative failures by inspection of the raw time series. Once the epidemic period was defined, the size of an epidemic per antigenic variant was calculated using the estimated resident population for that particular year and city.

**Normalisation of epidemic incidence.** For each epidemic, the incidence of laboratory-confirmed cases per million people was calculated from the number of raw counts. Given the positive skew in the distribution of epidemic incidences, individual incidence values were log-transformed. To enable comparisons within subtypes, we needed to account for potential differences in surveillance intensity, and normalise values between cities: we subtracted off the overall city-specific mean log-transformed incidence from each individual value. Although the apparent heterogeneity in the effect of antigenic change and prior immunity between subtypes suggests that data should be stratified by subtype, we repeated our analyses with data aggregated and normalised across subtypes in order to increase statistical power. Individual log-transformed values for each epidemic were instead transformed by subtracting off the overall city- and subtype-specific mean of the log-transformed values.

**Virus antigenic characterisation by haemagglutination inhibition assay.** For our analyses, we defined an antigenic variant as in Smith et al.<sup>14</sup>, where an antigenic variant is sufficiently different from preceding variants to warrant an update of the seasonal influenza virus vaccine. To this end, our analyses only accounted for major antigenic changes, and did not account for the possibility of small or gradual antigenic changes (neither of which are well studied for seasonal influenza viruses).

The haemagglutination inhibition (HI) assay data used in this study only compared the test virus and the then current reference vaccine strain to assess whether or not viruses had changed antigenically. However, this comparison to a single reference point is potentially problematic, given that new Southern Hemisphere’s influenza vaccine composition recommendations are made every September. This is usually after the end of the influenza season in Australia, and may lead to misidentification during antigenic characterisation of submitted samples during the preceding season where samples containing a novel antigenic variant may have been tested with sera raised against its predecessor variant. To ameliorate this potential source of bias, we compared the antigenic characterisation data against phylogenetic data. This comparison revealed two instances for A/H3 viruses where the reference strain comparison by HI was misleading regarding the antigenic composition of an epidemic. There were a substantial number of laboratory-confirmed cases attributable to A/H3/Fujian/411/2002-like viruses in 2004, but phylogenetic analyses of sequences dated 2004 show that the Fujian/411/2002-like viruses had already been replaced by the novel California/7/2004 variant viruses. Similarly, in 2005, a substantial number of samples initially identified as A/H3 California/7/2004-like viruses were phylogenetically in the new A/Wisconsin/67/2005 variant group.

To account for the likelihood of misidentification due to delays in updating nomenclature, we assumed that all A/H3 cases in 2004 were California/7/2004-like, and in 2005 were Wisconsin/67/2005-like antigenic variants. Additional analyses were also carried out with the raw data set without these corrections (see Supplementary Figs. 9–13 and Supplementary Table 6), and lead to no significant or substantive differences to our findings.

**Demographic data.** We retrieved estimated resident populations for Adelaide, Brisbane, Melbourne, Perth and Sydney on 30 June of each year from 2000 to 2015 from the Australian Bureau of Statistics (<http://stat.abs.gov.au/>).

**Climate data.** For each of the five cities, we compiled the mean temperature (°C) and relative humidity (%) from TuTiempo (<https://en.tutiempo.net/>), and calculated the mean absolute humidity ( $g\ m^{-3}$ ) for each 2-week period from 1985 to 2015. For each of the 26 2-week periods of the calendar year, we calculated 31-year mean temperature  $\bar{T}$  and absolute humidity  $\overline{AH}$  values (see Eqs. (5) and (6) below).

**Testing the statistical significance of anomalous absolute humidity and temperature.** Following the method presented by Shaman et al.<sup>10</sup>, we calculated local anomalous  $T'$  and absolute humidity  $AH'$  values for each city, and 2-week period of the year from 2000 to 2015. For each 2-week period,  $T'$  and  $AH'$  are

defined as the deviation in observed temperature  $T$  and actual absolute humidity  $AH$  from their 31-year mean values,  $\bar{T}$  and  $\overline{AH}$ , respectively (Eqs. (5) and (6)).

$$T' = T - \bar{T} \quad (5)$$

$$AH' = AH - \overline{AH} \quad (6)$$

Following Shaman et al.<sup>10</sup>, we generated a synthetic distribution of wintertime climatic values by bootstrap sampling. In order to maintain the sampling structure and control for anomaly variability among the cities, 15 n-week continuous blocks were randomly sampled from 1 April–31 August, 1985–2015 for each of the five cities. These 75 samples were then averaged to produce a mean  $T'$  and  $AH'$  value. This was repeated 100,000 times to produce a bootstrapped distribution of average values. The statistical significance for the mean  $T'$  and  $AH'$  values derived from the 75 empirically observed earliest-in-the-season epidemics was then calculated non-parametrically, by determining the quantile for the observed values within the bootstrap distributions. This bootstrap was repeated at the city level to see if there were geographical differences with individual bootstrap distributions that were created for each city.

We also evaluated whether or not epidemic onset is associated with climatic fluctuations that are anomalous for that particular time of the year. By definition, for any given 2-week period of the year, the 31-year mean for  $T'$  and  $AH'$  is 0. We used a Wilcoxon one-sample test to assess whether there were reductions in climatic values in the observed set of  $T'$  and  $AH'$  values in each of the 2-week blocks preceding the onset of the earliest epidemic of the season.

**Bayesian hierarchical regression.** To estimate reasonable bounds on the possible effects of climate and antigenicity on epidemic size, we used a Bayesian hierarchical model that partially pooled effect-size estimates across subtypes, increasing the capacity to detect any potential effects without assuming a priori that effects should be the same across different subtypes. We fit the model using Markov Chain Monte Carlo (MCMC) with Stan<sup>54</sup> and its R interface rstan<sup>55</sup>; Stan implements a no-turn sampler (NUTS)<sup>54</sup>. All data and code needed to reproduce the analysis and figures are provided in the project Github repository, along with directions in a README file.

In the model notation that follows, the symbol “ $\sim$ ” is a “sampling statement”; it denotes that a random variable is distributed according to the given distribution. Normal distributions are parameterised as Normal(mean, standard deviation), generalised Student-T distributions are parameterised as Student-T(degrees of freedom, location and scale). Positive-constrained normal distributions (Half-Normal) are parameterised as Half-Normal(mode, standard deviation).

We predicted log incidence minus city- and subtype-specific mean log incidence as a function of the following predictor variables:

- $X_1$ : whether the epidemic was the first epidemic for an antigenic variant in the city (binary, yes or no)
- $X_2$ : cumulative prior incidence of the antigenic variant (measured as log(total prior cases/city- and subtype-specific mean cases per epidemic))
- $X_3$ : mean absolute humidity during the epidemic, from the start to end date of the epidemic (measured as fortnight of the year)
- $X_4$ : start date of the epidemic (measured as fortnight of the year)
- $X_5$ : whether the epidemic was the earliest epidemic (of any subtype) in the city that year (binary, yes or no)
- $X_6$ : the cumulative amount of influenza activity (of any subtype) in the city that year prior to the epidemic
- $X_7$ : mean rainfall during the epidemic, from the start to end date of the epidemic (measured as fortnight of the year).

We omitted mean epidemic temperature as a predictor as it was highly collinear with absolute humidity. Any observed large effect of absolute humidity could therefore theoretically have been attributable to temperature, though in practice we estimated an effect near zero for absolute humidity.

We made a linear prediction of an epidemic's normalised size given its values for  $X = (X_1, \dots, X_7)$ . Effect sizes  $b_i$  for each predictor  $X_i$  were subtype-specific, with  $b_{ij}$  denoting the effect of variable  $i$  for subtype  $j$ . We also estimated subtype-specific intercepts  $a_j$ .

We included cumulative antigenic variant activity and prior activity in the year only for old antigenic variants and epidemics that were not first of the year, respectively, that is, as interaction terms with one minus the corresponding binary variables. So the predicted mean-centred log size  $\langle y_k \rangle$  of an epidemic of subtype  $j$  is given by Eq. (7), where  $X_{ik}$  denotes the value of  $X_i$  for epidemic  $k$ . Following Gelman<sup>51</sup>, we mean-centred and scaled continuous predictors so that effect sizes  $b$  would be directly comparable between binary and continuous predictors.

$$\langle y_k \rangle = a_j + b_{1j}X_{1k} + b_{2j}X_{2k}(1 - X_{1k}) + b_{3j}X_{3k} + b_{4j}X_{4k} + b_{5j}X_{5k} + b_{6j}X_{6k}(1 - X_{5k}) + b_{7j}X_{7k} \quad (7)$$

We assumed that observed normalized log epidemic sizes  $y_k$  were normally distributed about their predicted log sizes  $\langle y_k \rangle$  with an unknown, estimated standard deviation  $\sigma_y$  (Eq. (8)):

$$y_k \sim \text{Normal}(\langle y_k \rangle, \sigma_y) \quad (8)$$

We assumed that subtype effect sizes  $b_{ij}$  for each predictor  $i$  and subtype  $j$  were normally distributed about a general mean effect size  $\langle b_i \rangle$ , with an unknown, estimated predictor-specific standard deviation  $\sigma_{bi}$  (Eq. (9)):

$$b_{ij} \sim \text{Normal}(\langle b_i \rangle, \sigma_{bi}) \quad (9)$$

Likewise, we assumed that intercepts  $a_i$  were normally distributed about a mean intercept  $\langle a \rangle$  with an unknown, estimated standard deviation  $\sigma_a$  (Eq. (10)):

$$a_i \sim \text{Normal}(\langle a \rangle, \sigma_a) \quad (10)$$

We assumed that predictor-specific effect-size standard deviations  $\sigma_{bi}$  were half-normally distributed with mode 0 and an unknown, estimated standard deviation  $\sigma_b$  (Eq. (11)).

$$\sigma_{bi} \sim \text{Half} - \text{Normal}(0, \sigma_b) \quad (11)$$

We placed weakly informative<sup>56</sup> positive-constrained half-normal priors on the intercept, effect size and error-term standard deviations  $\sigma_a$ ,  $\sigma_b$  and  $\sigma_y$  (Eqs. (12–14)). Weakly informative priors rule out biologically or mathematically implausible parameter values while allowing data rather than assumptions to inform inferences regarding plausible values.

$$\sigma_a \sim \text{Half} - \text{Normal}(0, 0.5) \quad (12)$$

$$\sigma_b \sim \text{Half} - \text{Normal}(0, 1) \quad (13)$$

$$\sigma_y \sim \text{Half} - \text{Normal}(0, 1) \quad (14)$$

We placed a weakly informative Gaussian prior on the mean intercept  $\langle a \rangle$  (Eq. (15)) and a weakly informative Student-T prior on the mean effect sizes  $\langle b_i \rangle$  (Eq. (16)):

$$\langle a \rangle \sim \text{Normal}(0, 1) \quad (15)$$

$$\langle b_i \rangle \sim \text{Student} - T(3, 0, 2.5) \quad (16)$$

The intercept prior was based on the degree of variation in the normed outcome variable to cover it while ruling out intercepts much larger or smaller than the largest and smallest observations. The effect-size prior was based on a recommendation for weakly informative regression effect-size priors (for scaled predictors) from the Stan prior recommendation wiki (<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>).

We ran four MCMC chains, each with a 1000-step sample warmup period followed by 1000 saved posterior samples, for a total of 4000 posterior draws. We verified convergence by inspecting trace plots, and confirming that all parameters had sufficiently low Rhat values (all Rhat < 1.005) and sufficiently large effective sample sizes (all neff > 16% of total sample size). We visualised posteriors as quantile dotplots<sup>57</sup> to aid in visual estimation of distributions.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All of the data for these statistical analyses and models are available at the following Github repository: [https://github.com/edwardkslam/australian\\_seasonal\\_flu](https://github.com/edwardkslam/australian_seasonal_flu).

## Code availability

Code developed for these statistical analyses and models is available at the following Github repository: [https://github.com/edwardkslam/australian\\_seasonal\\_flu](https://github.com/edwardkslam/australian_seasonal_flu).

Received: 15 August 2019; Accepted: 9 May 2020;

Published online: 02 June 2020

## References

- World Health Organization. Influenza (Seasonal) Fact sheet. [https://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)) (2018).
- Denoeud, L. et al. Predicting pneumonia and influenza mortality from morbidity data. *PLoS ONE* **2**, 3–6 (2007).
- Finkelman, B. S. et al. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: Viral coexistence and latitudinal gradients. *PLoS One* **2**, e1296 (2007).
- Gandon, S., Day, T., Metcalf, C. J. E. & Grenfell, B. T. Forecasting epidemiological and evolutionary dynamics of infectious diseases. *Trends Ecol. Evol.* **31**, 776–788 (2016).
- Public Health England. Flu plan winter 2016/17. *Public Heal. Engl.* <https://doi.org/10.1037/e500942012-001> (2016).

6. Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **3**, 1470–1476 (2007).
7. Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc. Natl Acad. Sci. USA* **106**, 3243–3248 (2009).
8. Lowen, A. C. & Steel, J. Roles of humidity and temperature in shaping influenza seasonality. *J. Virol.* **88**, 7692–7695 (2014).
9. Tamerius, J. D. et al. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathog.* **9**, e1003194 (2013).
10. Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316 (2010).
11. Shoji, M., Katayama, K. & Sano, K. Absolute humidity as a deterministic factor affecting seasonal influenza epidemics in Japan. *Tohoku J. Exp. Med.* **224**, 251–256 (2011).
12. Davis, R. E., Rossier, C. E. & Enfield, K. B. The impact of weather on influenza and pneumonia mortality in New York City, 1975–2002: a retrospective study. *PLoS ONE* **7**, e34091 (2012).
13. Seleznev, N. E. & Leonenko, V. N. Absolute humidity anomalies and the influenza onsets in Russia: a computational study. *Procedia Comput. Sci.* **119**, 224–233 (2017).
14. Smith, D. J. Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376 (2004).
15. Langat, P. et al. Genome-wide evolutionary dynamics of influenza B viruses on a global scale. *PLoS Pathog.* **13**, e1006749 (2017).
16. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of inter-pandemic influenza A (H3N2) in humans. *Science* **6829**, 1898–1904 (2006).
17. Koelle, K., Kamradt, M. & Pascual, M. Understanding the dynamics of rapidly evolving pathogens through modeling the tempo of antigenic change: influenza as a case study. *Epidemics* **1**, 129–137 (2009).
18. Bock Axelsen, J., Yaari, R., Grenfell, B. T. & Stone, L. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proc. Natl Acad. Sci. USA* **111**, 9538–9542 (2014).
19. Greene, S. K., Ionides, E. L. & Wilson, M. L. Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968–1998. *Am. J. Epidemiol.* **163**, 316–326 (2006).
20. Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
21. Charu, V. et al. Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput. Biol.* **13**, 1–23 (2017).
22. Sakai, T. et al. Geographic and temporal trends in influenzalike illness, Japan, 1992–1999. *Emerg. Infect. Dis.* **10**, 1822–1826 (2004).
23. Geoghegan, J. L. et al. Continental synchronicity of human influenza virus epidemics despite climatic variation. *PLoS Pathog.* **14**, 1–16 (2018).
24. Schnepf, N. et al. High burden of non-influenza viruses in influenza-like illness in the early weeks of H1N1v epidemic in France. *PLoS ONE* **6**, e23514 (2011).
25. Australian Bureau of Statistics. ERP by SA2 and above (ASGS 2001), 1991 to 2016. <http://www.abs.gov.au> (2017).
26. Tamerius, J. et al. Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environ. Health Perspect.* **119**, 439–445 (2011).
27. Chattopadhyay, I., Kiciman, E., Elliott, J. W. & Shaman, J. L. Conjunction of factors triggering waves of seasonal influenza. *eLife* **7**, e30756 (2018).
28. Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S. & Sugihara, G. Global environmental drivers of influenza. *Proc. Natl Acad. Sci. USA* **113**, 13081–13086 (2016).
29. Kucharski, A. J., Andreasen, V. & Gog, J. R. Capturing the dynamics of pathogens with many strains. *J. Math. Biol.* **72**, 1–24 (2016).
30. Wu, A., Peng, Y., Du, X., Shu, Y. & Jiang, T. Correlation of influenza virus excess mortality with antigenic variation: application to rapid estimation of influenza mortality burden. *PLoS Comput. Biol.* **6**, e1000882 (2010).
31. Bedford, T. et al. Integrating influenza antigenic dynamics with molecular evolution. *eLife* **3**, 1–26 (2014).
32. Yang, W., Lipsitch, M. & Shaman, J. Inference of seasonal and pandemic influenza transmission dynamics. *PLoS Natl Acad. Sci. USA* **112**, 2723–2729 (2015).
33. Barr, I. G. et al. A new pandemic influenza A(H1N1) genetic variant predominated in the winter 2010 influenza season in Australia, New Zealand and Singapore. *Eurosurveillance* **15**, 1–6 (2010).
34. Hoschler, K. et al. Seroprevalence of influenza A (H1N1) pdm09 Virus. *Emerg. Infect. Dis.* **18**, 2010–2013 (2012).
35. Lessler, J. et al. Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. *PLoS Pathog.* **8**, 26 (2012).
36. Kucharski, A. J. et al. Estimating the life course of influenza A(H3N2) antibody responses from cross-sectional data. *PLoS Biol.* **13**, 1–16 (2015).
37. Neuzil, K. M. M. et al. Immunogenicity and reactogenicity of 1 versus 2 doses of trivalent inactivated influenza vaccine in vaccine-naïve 5–8-year-old children. *J. Infect. Dis.* **194**, 1032–1039 (2006).
38. Gouma, S., Kim, K., Weirick, M., Gumina, M. E. & Branche, A. Middle-aged individuals may be in a perpetual state of H3N2 influenza virus susceptibility. Preprint at <https://www.medrxiv.org/content/10.1101/2020.01.09.20017038v1.full.pdf> (2020).
39. Meade, P. D. et al. Influenza virus infection induces a narrow antibody response in children but a broad recall response in adults. *mBio* **11**, 1–15 (2020).
40. Lee, J. M. et al. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife* **8**, e49324 (2019).
41. Dalziel, B. D. et al. Contact heterogeneity, rather than transmission efficiency, limits the emergence and spread of canine influenza virus. *PLoS Pathog.* **10**, e1004455 (2014).
42. Dalziel, B. D. et al. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science* **362**, 75–79 (2018).
43. Sonoguchi, T., Naito, H., Hara, M., Takeuchi, Y. & Fukumi, H. Cross-subtype protection in humans during sequential, overlapping, and/or concurrent epidemics caused by h3n2 and h1n1 influenza viruses. *J. Infect. Dis.* **151**, 81–88 (1985).
44. Cowling, B. J. et al. Protective efficacy of seasonal influenza vaccination against seasonal and pandemic influenza virus infection during 2009 in Hong Kong. *Clin. Infect. Dis.* **51**, 1370–1379 (2010).
45. Goldstein, E., Cobey, S., Takahashi, S., Miller, J. C. & Lipsitch, M. Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. *PLoS Med.* **8**, 1–12 (2011).
46. Bedford, T. et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523**, 217–220 (2015).
47. Thompson, W. W. et al. Mortality associated with influenza and respiratory syncytial virus in the United States. *Am. Med. Assoc.* **289**, 179–186 (2003).
48. Caini, S., Kroneman, M., Wieggers, T., El Guerche-Séblain, C. & Paget, J. Clinical characteristics and severity of influenza infections by virus type, subtype, and lineage: a systematic literature review. *Influenza Other Respir. Viruses* **12**, 780–792 (2018).
49. Barr, I. G., Vijaykrishna, D. & Sullivan, S. G. Differential age susceptibility to influenza B/victoria lineage viruses in the 2015 Australian influenza season. *Eurosurveillance* **21**, 1–9 (2016).
50. Sullivan, S. G., Chilver, M. B. N., Higgins, G., Cheng, A. C. & Stocks, N. P. Influenza vaccine effectiveness in Australia: results from the Australian Sentinel Practices Research Network. *Med. J. Aust.* **201**, 109–111 (2014).
51. Höhle, M., Riebler, A. & Paul, M. Getting started with outbreak detection. <https://cran.r-project.org/web/packages/surveillance/vignettes/surveillance.pdf> (2007).
52. Salmon, M., Schumacher, D. & Höhle, M. Monitoring count time series in R: aberration detection in public health surveillance. *J. Stat. Softw.* **70**, 10 (2016).
53. Velleman, P. F. Definition and comparison of robust nonlinear data smoothing algorithms. *J. Am. Stat. Assoc.* **75**, 609 (1980).
54. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual. <https://mc-stan.org/users/documentation/> (2018).
55. Stan Development Team. RStan: the R interface to Stan. <http://mc-stan.org/> (2018).
56. Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y.-S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**, 1360–1383 (2008).
57. Fernandes, M., Walls, L., Munson, S., Hullman, J. & Kay, M. Uncertainty displays using quantile dotplots or CDFs improve transit decision-making. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems* 1–12, <https://doi.org/10.1145/3173574.3173718> (ACM Press, 2018).

## Acknowledgements

The authors thank the WHO Collaborating Centre for Reference and Research on Influenza (Victorian Infectious Diseases Reference Laboratory, Melbourne, Australia), which is supported by the Australian Government Department of Health, for the virological analysis of samples submitted by sentinel health practitioners and surveillance systems.

## Author contributions

E.K.S.L., D.H.M. and C.A.R. designed the research; E.K.S.L. and D.H.M. performed the data analyses; A.C.H. and I.G.B. generated the influenza virus data; E.K.S.L., D.H.M. and C.A.R. wrote the first draft of the paper. All authors contributed to the critical review and revision of the paper.

## Competing interests

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-16545-6>.

**Correspondence** and requests for materials should be addressed to C.A.R.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020