bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Accounting for healthcare-seeking behaviours and testing practices in real-time influenza forecasts

Robert Moss [1]*, Alexander E Zarebski [2], Sandra J Carlson [3], James M McCaw [1,4,5,6]

[1] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Australia.
[2] Department of Zoology, The University of Oxford, Oxford, England.
[3] Hunter New England Population Health, Newcastle, Australia.
[4] School of Mathematics and Statistics, The University of Melbourne, Parkville, Australia.
[5] Murdoch Children's Research Institute, The Royal Children's Hospital, Parkville, Australia.
[6] Victorian Infectious Diseases Reference Laboratory Epidemiology Unit, Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Parkville, Australia.
* Correspondence: rgmoss@unimelb.edu.au; Tel.: +61-3-8344-9430

Abstract: For diseases such as influenza, where the majority of infected persons experience mild (if any) symptoms, surveillance systems are sensitive to changes in healthcare-seeking and clinical decision-making behaviours. This presents a challenge when trying to interpret surveillance data in near-real-time (e.g., in order to provide public health decision-support). Australia experienced a particularly large and severe influenza season in 2017, perhaps in part due to (a) mild cases being more likely to seek healthcare; and (b) clinicians being more likely to collect specimens for RT-PCR influenza tests. In this study we used weekly Flutracking surveillance data to estimate the probability that a person with influenza-like illness (ILI) would seek healthcare and have a specimen collected. We then used this estimated probability to calibrate near-real-time seasonal influenza forecasts at each week of the 2017 season, to see whether predictive skill could be improved. While the number of self-reported influenza tests in the weekly surveys are typically very low, we were able to detect a substantial change in healthcare seeking behaviour and clinician testing behaviour prior to the high epidemic peak. Adjusting for these changes in behaviour in the forecasting framework improved predictive skill. Our analysis demonstrates a unique value of community-level surveillance systems, such as Flutracking, when interpreting traditional surveillance data.

Keywords: influenza; epidemics; forecasting; public health; surveillance; ascertainment

1. Introduction

Public health surveillance systems provide valuable insights into disease incidence, which can inform preparedness and response activities [1]. The data obtained from these systems must be interpreted appropriately, which requires an understanding of the characteristics of each system. For example, there is a well-established need to "understand the processes that determine how persons are identified by surveillance systems in order to appropriately adjust for the biases that may be present" [1]. The importance of understanding how people are identified by a surveillance system is particularly problematic in the context of near-real-time influenza forecasting [2–10], because healthcare-seeking behaviours and clinical decision-making are dynamic [11,12] and are subject to acute influences (e.g., media coverage [13]). The challenge that this poses is further compounded by delays in data collection and reporting, which reduce forecast performance [9,14,15]. And in the event of a pandemic, changes in patient and clinician behaviours are likely to be even more pronounced, with ramifications not only for interpreting surveillance data [3], but also for delivering effective and proportionate public health responses [16]. For influenza and influenza-like illnesses (ILI), the Flutracking online community surveillance system offers unique perspectives on disease incidence and severity in Australia, but also on healthcare-seeking behaviours and clinical decision-making [17–20].

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The 2017 influenza season was particularly severe in most of Australia, with an unprecedented number of laboratory-confirmed influenza cases, and high consultation, hospitalisation, and mortality rates [21]. The 2017 season presented a very real challenge to public health staff in Australia for a number of reasons, among which were the unprecedented scale of the surveillance data and the delays in reporting that this entailed. In the subsequent 2017/18 northern hemisphere influenza season, the dominant strain was frequently referred to as the "Australian flu" [22].

We have previously adapted epidemic forecasting methods to the Australian context [23]. Notable features of this context include: more than half of the population reside in the five largest cities (Sydney, Melbourne, Brisbane, Perth, Adelaide), which are sparsely distributed across a landmass approximately the same size as continental USA; ILI data are available from sentinel surveillance systems and public health services, but on a much smaller scale than, e.g, the ILINet system in the USA; and laboratory-confirmed influenza infections data are available, but testing denominators are typically only provided by public laboratories (which comprise only a small proportion of all laboratory testing). In previous work we have applied these forecasting methods to individual Australian cities, and have used them to compare multiple surveillance systems [3], developed model-selection methods to measure the benefit of incorporating climatic effects [24], and have deployed these forecasts in real-time in collaboration with public health staff [9,10]. Most recently, we have used these methods to quantify the impact of delays in reporting and to highlight challenges in accounting for population and clinician behaviour changes in response to a severe season [10].

In 2017 we produced weekly near-real-time influenza forecasts for the capital cities of three Australian states: Sydney (New South Wales), Brisbane (Queensland), and Melbourne (Victoria). As described previously, we performed a one-off re-calibration of our real-time influenza forecasts (2–5 weeks prior to the observed peaks), which improved the forecast performance [10]. This was a manual process and was informed by qualitative expert opinion rather than by quantitative methods. In the 2018 influenza season we also began producing weekly near-real-time influenza forecasts for Perth, the capital city of Western Australia.

Here, we investigate how the use of community-level ILI surveillance data, as captured by the Flutracking surveillance system [17–20], may allow us to re-calibrate our forecasts in real-time based on quantitative measures, and potentially improve forecast performance. We use the 2014–16 seasons in these four cities for calibration, and the 2017 season in these four cities as a case study.

2. Results

Retrospective analysis revealed that in the 2017 influenza season there was a marked increase in the probability that Flutracking participants with ILI symptoms (defined in Methods section 4.1) would seek healthcare and have a specimen collected for testing, relative to previous influenza seasons, in all states except Western Australia (Figure 1). In the other influenza seasons (2014–2016) there was no such increase; this probability remained very similar to the average of the 2014 season (horizontal dashed lines in Figure 1). Queensland was the one exception, where this probability did increase in 2015 and 2016, but not by the same scale as observed in 2017.

Consistent with the Flutracking analysis shown in Figure 1, the 2017 influenza season was substantially larger than previous influenza seasons in all cities except Perth (Figure 2).

2.1. Relating Flutracking trends to the general population in 2014–16

We hypothesised that these trends in influenza testing would also be reflected in the general population, but that there may be a *difference in scale* between the behaviour changes reported in the Flutracking surveys and those of the general population. We introduced a scaling factor κ to characterise the relationship between the testing probability obtained from the Flutracking surveys and the testing probability of the general population (Figure 3). When $\kappa = 1$ the change in the population testing probability $p_{\text{id}}(w)$ — relative to its initial value $p_{\text{id}}(0)$ — is the same as the change in the Flutracking testing probability $P\,(\text{Test}|\text{ILI}, \textit{week} = w)$ relative to its initial value

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

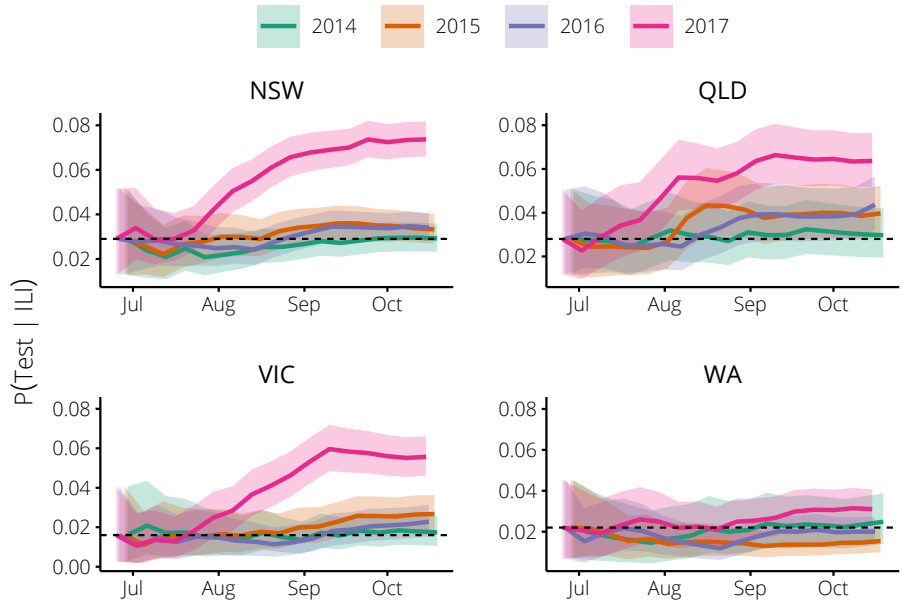

Figure 1. The probability of being swabbed for an influenza test, given ILI symptoms, as estimated from the weekly Flutracking survey data in each of the 2014–2017 seasons for participants in New South Wales (top-left), Queensland (top-right), Victoria (bottom-left), and Western Australia (bottom-right). The horizontal dashed lines indicate the prior expectation for each state, the solid lines indicate the posterior expectation, after applying our statistical model, and the shaded regions indicate the 95% median-centred credible interval. In all states except Western Australia, this probability was much higher in 2017 than in previous years. Queensland also exhibited a moderate increase in 2015.

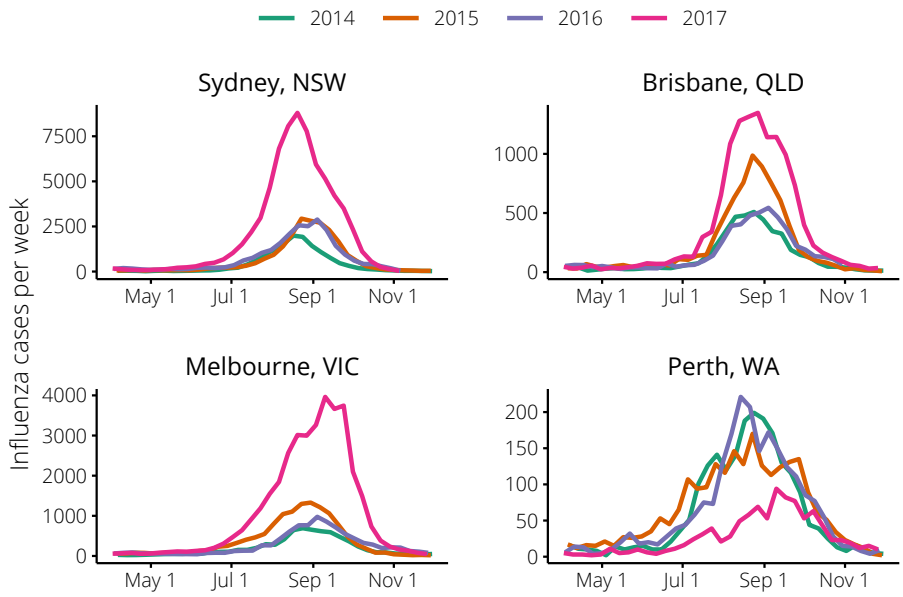

Figure 2. The weekly influenza surveillance data in each of the 2014–2017 seasons for the cities of: Sydney, New South Wales (top-left); Brisbane, Queensland (top-right); Melbourne, Victoria (bottom-left); and Perth, Western Australia (bottom-right). In all states except Western Australia, the 2017 season was much larger than the 2014–2016 seasons. Queensland also experienced quite a large season in 2015.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

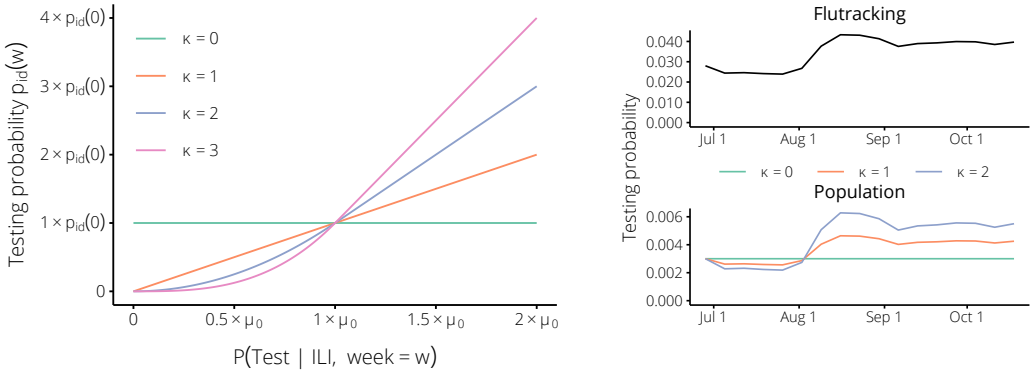

Figure 3. The testing probability $p_{\mathrm{id}}(w)$ for the general population is a smooth function of the testing probability $P\,(\mathrm{Test}|\mathrm{ILI}, week = w)$ obtained from the Flutracking data (left panel). Note that this relationship is non-linear as the Flutracking testing probability approaches zero, in order to ensure that the population testing probability remains non-negative, and that the Flutracking testing probability is defined in terms of its initial value $\mu_0 = P\,(\mathrm{Test}|\mathrm{ILI}, week = 0)$. As the Flutracking testing probability changes (right panel, top), the population testing probability will also change (right panel, bottom).

$\mu_0 = P\,(\mathrm{Test}|\mathrm{ILI}, week = 0)$. When $\kappa > 1$, the population testing probability is more sensitive to changes in the Flutracking testing probability. And when $\kappa = 0$ we recover the *non-calibrated forecasts* — i.e., the forecasts obtained without accounting for any behaviour changes. We used retrospective forecasts to identify an optimal value for this scaling parameter.

We produced retrospective weekly influenza forecasts for the 2014, 2015, and 2016 influenza seasons for the capital cities of four Australian states: Sydney (New South Wales), Brisbane (Queensland), Melbourne (Victoria), and Perth (Western Australia). At each week, the forecasts were calibrated based on the Flutracking survey data, by adjusting the *testing probability* $p_{\mathrm{id}}(w)$ at each week w. This testing probability was used to relate disease incidence in epidemic simulations to reported surveillance data (see Methods section 4.4). We evaluated how this calibration affected forecast performance, according to four different forecasting targets:

- "Now-casts": how well forecasts predicted the observation for the forecasting date itself.
- "Next 2 weeks": how well forecasts predicted observations 1–2 weeks after the forecasting date.
- "Next 4 weeks": how well forecasts predicted observations 1–4 weeks after the forecasting date.
- "Next 6 weeks": how well forecasts predicted observations 1–6 weeks after the forecasting date.

Forecast performance was evaluated using Bayes factors, which are the ratio of the likelihood that the available data support two competing hypotheses [24,25]. In this study we used the ratio of the calibrated forecast likelihood (where $\kappa = k$) to the non-calibrated forecast likelihood, averaged over all of the forecasting weeks, and reported the logarithm $\widetilde{b_k}$ of this average, which we denote the "mean performance" (see Methods section 4.5 for details). Where $\widetilde{b_k} > 1$ the *calibrated forecasts* are considered to be more strongly supported by the data, and where $\widetilde{b_k} < -1$ the *non-calibrated forecasts* are considered to be more strongly supported. To measure forecast performance over multiple years, we normalised the mean performance $\widetilde{b_k}$ in each year and reported the mean $\widehat{b_k}$ of these annual values, which we denote the "mean normalised performance".

There were consistent trends across these four targets in each city, and substantial differences in performance between the cities, as summarised in Figure 4. For values of κ between 1 and 5, the calibrated forecasts strongly out-performed the non-calibrated forecasts in both Brisbane and Melbourne. Over a similar range, the calibrated forecasts showed no improvement in Sydney, and performed much worse in Perth than the non-calibrated forecasts. Note that in Sydney, for $0 < \kappa \leq 2$ the forecast performance was not meaningfully better *or* worse than the non-calibrated forecasts ($\kappa = 0$). For larger values of κ, the calibrated forecasts were out-performed by the non-calibrated forecasts in

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

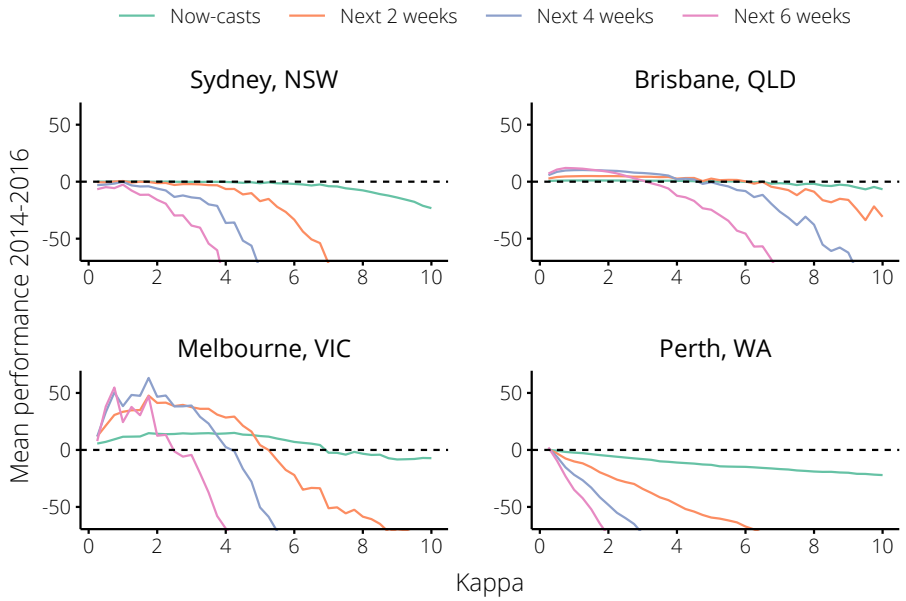

Figure 4. Mean forecast performance $\widetilde{b_k}$ over the 2014–16 influenza seasons, relative to the non-calibrated forecasts (the horizontal dashed line); $\widetilde{b_k} > 1$ indicates substantial performance improvement, and $\widetilde{b_k} < -1$ indicates substantial performance reduction. Performance was measured against four different targets (see text for details), over a range of values for the scaling parameter κ. For small values of κ, forecast calibration significantly increased skill in Brisbane and Melbourne, had little effect in Sydney, and decreased skill in Perth.

all cities and for all targets. Based on the mean *normalised* performance in all cities *except* Perth, we identified an optimal value $\kappa = 1.75$ (Figure 5). Note that this value is greater than 1, which suggests that the behaviour of the general population exhibits larger changes than that of the Flutracking cohort.

2.2. Incorporating Flutracking trends into 2017 forecasts

The next question was whether the optimal value of κ, as chosen based on the results for 2014–16, would have led to improved forecast performance in 2017. In other words, are these findings of value in a real-time context?

In a retrospective analysis, we measured the forecast performance for the 2017 influenza season over the same range of values for κ, in order to evaluate how well the chosen value of $\kappa = 1.75$ would have performed. Note that this represents what would have been possible in real-time in the 2017 season. The calibrated testing probabilities $p_{\mathrm{id}}(w)$ are shown in Figure 6, and the resulting performance according to each of the forecasting targets is shown in Figure 7.

When we set $\kappa = 1.75$ (the optimal value according to the 2014–16 forecasts) forecast performance in 2017 was near-optimal according to all forecasting targets in all cities except Perth (Figure 7). This was even true of Sydney, where similar values of κ in the 2014–16 seasons had minimal impact on forecast performance. Note that the relative magnitudes of the performance according to each target is to be expected, since each target considers a different number of observations, and so a ten-fold improvement in predicting a single observation causes a hundred-fold improvement in predicting two observations.

For illustration, a subset of the Brisbane 2017 forecasts are shown in Figure 8, which shows how well the forecast predictions match the reported data over a range of values for κ, and at different times during the 2017 season.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

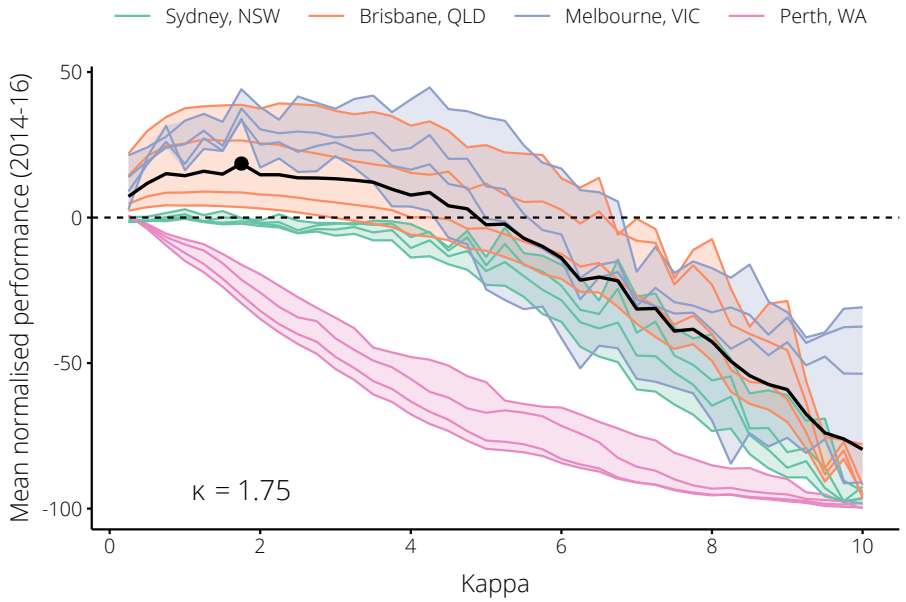

Figure 5. The trends in mean annually-normalised performance $\widehat{b}_k$ for each city over the 2014–16 influenza seasons, relative to the non-calibrated forecasts (the horizontal dashed line). Each set of coloured lines depicts the performance $\widehat{b}_k$ for a city against the four chosen forecasting targets (see text for details). The solid black line is the overall trend for all cities *except* Perth; the optimal value of κ according to this trend is indicated by the black circle on this line ($\kappa = 1.75$).

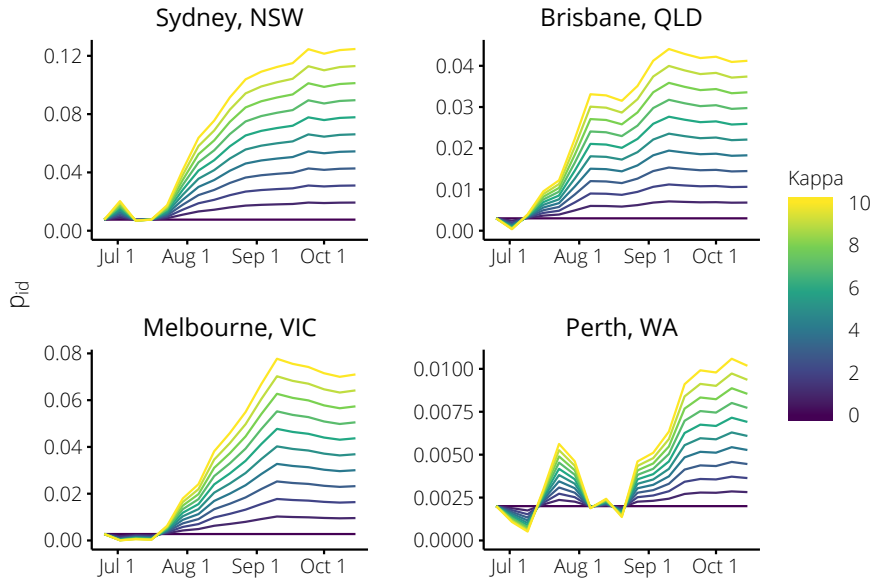

Figure 6. The testing probabilities $p_{\text{id}}(w)$ for each value of κ in each city during the 2017 season. When $\kappa = 1$ the (relative) change in $p_{\text{id}}(w)$ is the same as the (relative) change in the Flutracking testing probability; when $\kappa > 1$ $p_{\text{id}}(w)$ is more sensitive to changes in the Flutracking testing probability; and when $\kappa = 0$ we recover the *non-calibrated forecasts* (which do not account for any behaviour changes).

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

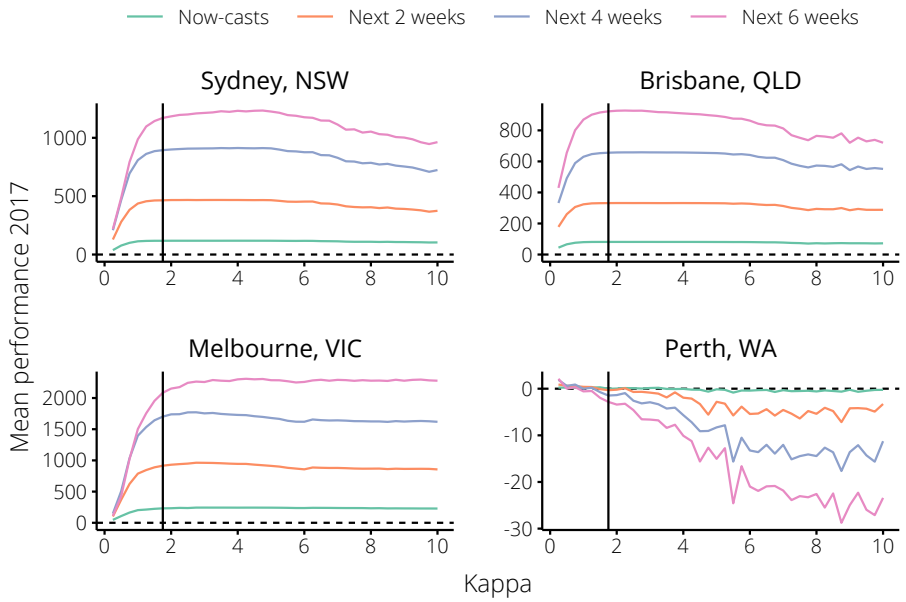

Figure 7. The trends in mean forecast performance $\widetilde{b}_k$ over the 2017 influenza season for each city, relative to the non-calibrated forecasts (the horizontal dashed line); $\widetilde{b}_k > 1$ indicates substantial performance improvement, and $\widetilde{b}_k < -1$ indicates substantial performance reduction. The black vertical line indicates $\kappa = 1.75$, the value chosen as optimal based on the 2014–16 seasons.

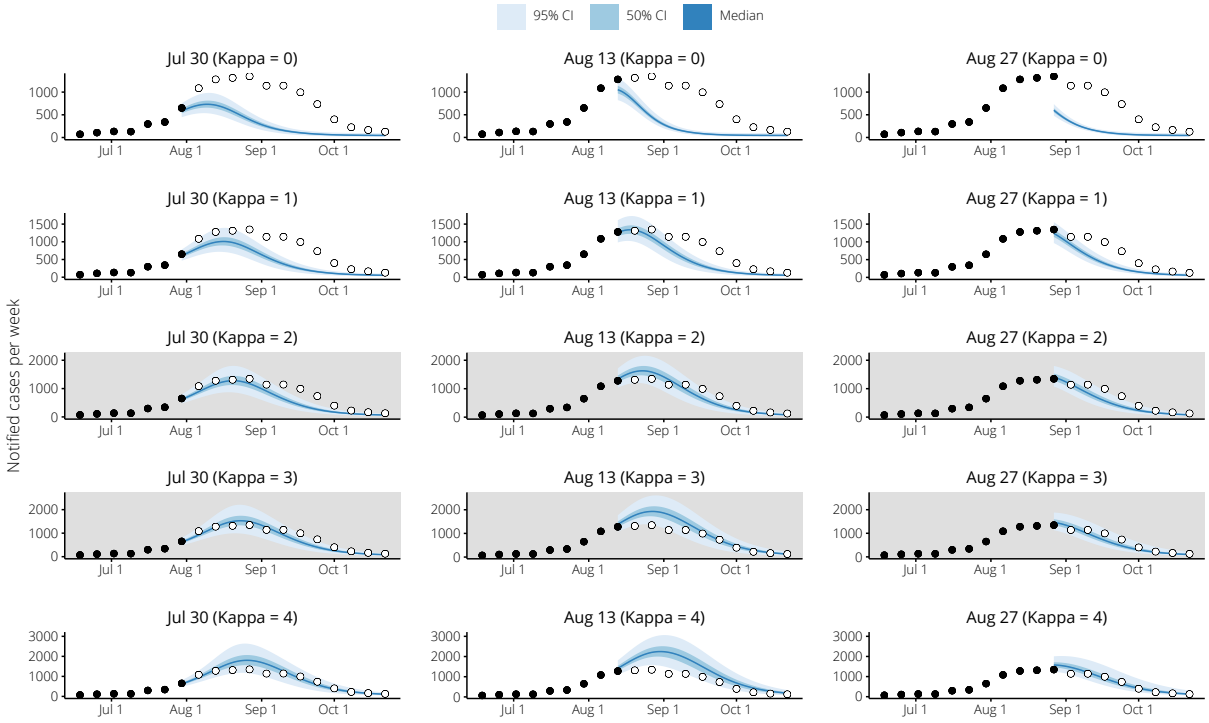

Figure 8. Example 2017 forecasts for Brisbane, Queensland, for various values of κ. Each column show forecasts generated for the same date, and each row shows forecasts generated using the same value of κ. The optimal values of κ for the Brisbane forecasting targets were 2–3 (shown with grey backgrounds), and this is supported by a visual inspection of the forecasts for each of the forecasting dates.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3. Discussion

3.1. Principal findings

Incorporating the behavioural trends estimated from the Flutracking self-reported influenza tests into the 2014–16 forecasts greatly improved influenza forecast performance in both Brisbane and Melbourne, but had negligible effect on performance in Sydney and substantially reduced performance in Perth (this was not unexpected, see the next section). We explored the relationship between the Flutracking survey data and the testing probability in the general population, identifying an optimal scaling value of $\kappa = 1.75$. This should be interpreted as meaning that behaviour changes in the general population were 75% larger than the behaviour changes characterised in the Flutracking survey data (where "behaviour changes" refers to healthcare-seeking behaviours and to clinician testing practices).

Using this same value for the scaling parameter in the 2017 forecasts greatly improved performance in both Brisbane and Melbourne, and also in Sydney. When we explored a wide range of values for the scaling parameter, these performance improvements were found to be near-optimal. Performance in Perth was reduced, but not as substantially as in the 2014–16 forecasts.

Similar performance improvements would have been obtained in the 2014–16 seasons, and in the 2017 season, if we had instead assumed that there was no difference between Flutracking participants and the general population (i.e., had we set $\kappa = 1$). This provides evidence that, with the possible exception of Perth, the behaviour changes characterised by the Flutracking data are not strongly biased.

These findings demonstrate that the insights into healthcare behaviours that are provided by community-level surveillance systems such as Flutracking can greatly improve the performance of influenza forecasting systems. This is particularly true in the event of an unusual epidemic, such as those observed in the 2017 Australian influenza season, where these behaviours may change substantially relative to previous seasons, and also over the course of the epidemic. Improvements in forecast performance are also likely in the event of a pandemic, where the perception of risk is likely to differ markedly from seasonal epidemics [26,27]. Incorporating these kinds of behavioural insights may also improve the performance of other infectious disease forecasting systems.

3.2. Study strengths and weaknesses

It is intrinsically difficult to incorporate the dynamics of human behaviour into an infectious disease modelling framework, as we have attempted here, because it involves many complexities and sources of uncertainty [11,28]. Retrospective analyses of data from multiple surveillance systems have previously been used to estimate disease severity in, e.g., the 2009 H1N1 pandemic in England [29], and evidence synthesis in a modelling framework is an active area of research [30]. However, the use of one surveillance system to calibrate the interpretation of another in real-time is, to our knowledge, a novel and significant contribution of this study.

The methods outlined here are also applicable beyond the Australian context. Surveillance systems similar to Flutracking operate around the world, including Influenzanet in many European countries [31] and Flu Near You in North America [32].

We have used data from multiple cities that routinely experience seasonal influenza epidemics, but which are geographically and climatically diverse. As the results for Perth show, incorporating behaviour changes into the forecasts is not guaranteed to improve forecast performance. But it does appear that the impact is reasonably consistent within a specific context, such as the same city from one year to the next. This was not an issue of sample size, since there are a similar number of participants from Western Australia as there are for other Australian states (see Appendix A). Instead, it is likely due to differences in the surveillance data (detailed in section 4.3). The Perth data were obtained from the public pathology laboratory (PathWest) and did not include serologic test results or data from private laboratories. Furthermore, these data contained a much higher proportion of hospitalised cases (40–60%) than did the data for the other cities in this study. Since people with severe disease will most

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

likely seek healthcare, regardless of behavioural influences, we would expect these data to be less sensitive to behaviour changes in the general community.

It may also reflect, in part, that Western Australia typically experiences an influenza season that is markedly different from the rest of the country [21]. The testing probability for Western Australia was near-constant and differed little across the 2014–17 seasons (Figure 1), and so it is possible that the "signal" in the Flutracking data for Western Australia was, in fact, noise. In contrast, for Sydney there was no clear signal in the Flutracking data for the 2014–16 seasons (Figure 1) and no improvement in forecast performance in these years (Figure 4). However, there was a marked change in the testing probability for Sydney in the 2017 season (Figure 1) and accounting for this yielded a substantial increase in forecast performance (Figure 7).

3.3. Meaning and implications

Infectious diseases that primarily cause mild (if any) symptoms are difficult to observe, especially in a *consistent manner*. The visibility of such cases to a surveillance system is sensitive to human behaviours that are primarily based on the perception of risk, which is subject to many influencing factors and can change rapidly [26,27]. Seasonal influenza is an archetypal example, which routinely presents a substantial challenge to healthcare systems in temperate climates, and whose clinical severity profile can differ substantially from one year to the next, and from setting to setting (e.g., city to city). Infectious disease forecasting is now beginning to establish itself as a useful decision-support tool for public health preparedness and response activities, particularly for seasonal influenza [2–10,14,15]. One of the many challenges for this field is to account for how behaviour changes affect surveillance data, and how these data should be interpreted [7–10]. Many sources of "non-traditional" data have been explored for their potential to improve infectious disease forecast performance, such as internet search engine queries and social media trends [33,34]. While these non-traditional data may indeed provide useful and valid insights, it is unclear how representative they are of disease burden [12,35], especially since those populations that experience the greatest burden of disease are also likely to be the most under-represented in these data [36]. Here, we have used a community-level surveillance system that is explicitly designed to capture influenza-like illness activity outside of healthcare systems, and associated healthcare-seeking behaviour. We argue that the Flutracking survey data provide a much more direct and meaningful quantification of how visible cases are to routine surveillance systems, and have shown here that incorporation of these data yield substantial improvements in forecast performance in what was an unprecedented influenza season.

During the early months of the 2017 Australian influenza season (up to the end of July), at which time we were not using Flutracking data to calibrate our forecasts [10], our near-real-time influenza forecasts were predicting the epidemic peaks to be similar in size to previous years, and to occur in early August. These would have been relatively early peaks — although not unusually so — and these predictions were *consistent with expert opinion at that time*. Had we been in a position to incorporate the Flutracking survey data into our forecasts at that time using the methods introduced in this study, we would have had advance warning that, contrary to our original predictions *and* expert opinion, we should have expected later (and, therefore, also larger) epidemic peaks. An early warning of this kind, which indicates that the current situation appears to be unusual and is inconsistent with "usual" expectations, is in itself a valuable decision-support tool, even in the absence of an accurate prediction of what will eventuate.

Seasonal influenza epidemics, and the threat of pandemic influenza, continue to present substantial challenges for public health preparedness and response in temperate climates, some 100 years after the 1918–19 influenza pandemic. Further complicating these activities are human behaviour changes, which confound the interpretation of surveillance data. However, as we have shown here, community-level surveillance is able to provide key insights into behaviour changes that can substantially improve near-real-time epidemic prediction and assessment. These advantages are of very real value to public health preparedness and response activities in the 21st century.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4. Materials and Methods

Access to the influenza surveillance data and the Flutracking survey data was approved by the Health Sciences Human Ethics Sub-Committee, Office for Research Ethics and Integrity, The University of Melbourne (Application ID 1646516.3).

4.1. Flutracking survey data

Flutracking weekly surveys were completed online voluntarily by community members in Australia and New Zealand. Participants responded to the weekly survey from April to October each year, and submitted surveys for themselves, and/or on behalf of other household members. The cohort of participants was maintained from year to year, unless participants unsubscribed. Participants were recruited online through workplace campaigns, current participants, the Flutracking website, traditional media and social media promotion.

Weekly counts of completed Flutracking surveys, self-reported ILI, and self-reported influenza tests were obtained from 2014 to 2017 for participants who had a usual postcode of residence in New South Wales, Victoria, Queensland, and Western Australia (see Appendix A for descriptive statistics of Flutracking data).

A participant with ILI was defined as having both self-reported fever and cough in the past week (ending Sunday). Any responses of "don't know" to "fever", "cough", or "time off work or normal duties" variables were removed from Flutracking data prior to data provision. This removed 0.9% of all surveys. Only new ILI symptoms were counted in the analysis. New symptoms were defined as the first week of a participant reporting ILI symptoms (where there were consecutive weeks of reporting ILI symptoms). If a participant reported ILI symptoms in one week, and then reported at least one week of no symptoms, followed by another report of symptoms, then this second symptom report was considered a new incident of ILI.

Participants with ILI who reported seeking health advice for their symptoms were asked whether they had been tested for influenza. Any self-reported influenza tests during an ILI incident were recorded in the same week of the new ILI symptoms (even if the test result was reported during the second, third or fourth week of illness). We acknowledge that Flutracking participant health seeking and testing behaviours may not be representative of the general Australian population. Flutracking has disproportionately high numbers of participants who are female, of working age, health care workers, have higher educational achievement, and have received influenza vaccination. Flutracking has a disproportionately lower number of participants who are Aboriginal or Torres Strait Islander [20].

4.2. Estimating the probability of an influenza test, given ILI symptoms

We used a Beta distribution to estimate the probability that a Flutracking participant with ILI symptoms sought healthcare and had a specimen collected for testing, because it is a suitable model for the distribution of a probability value. If we treat each participant with ILI symptoms as an independent binomial trial, where success is defined as that participant reporting they had a specimen collected for testing, then we obtain a Beta posterior distribution from a Beta prior distribution by conditioning on the results of each weekly Flutracking survey.

We specified separate prior means (μ_0) for each state, based on the 2014 Flutracking data for that state, and used a common prior variance (σ^2), in order to determine the initial shape parameters (α_0 and β_0):

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\sigma^2 = 10^{-4} \tag{1}$$

$$\alpha_0 = \mu_0 \cdot \left[\frac{\mu_0 \cdot (1-\mu_0)}{\sigma^2} - 1\right] \tag{2}$$

$$\beta_0 = (1-\mu_0) \cdot \left[\frac{\mu_0 \cdot (1-\mu_0)}{\sigma^2} - 1\right] \tag{3}$$

$$P(\text{Test}|\text{ILI}, week = 0) \sim \text{Beta}(\alpha_0, \beta_0) \tag{4}$$

$$\mathbb{E}\left[\text{Test}|\text{ILI}, week = 0\right] = \mu_0 \tag{5}$$

With N_w ILI cases and T_w self-reported tests in the Flutracking survey data for week w, we obtain the Beta posterior distribution for the probability that a Flutracking participant with ILI symptoms would seek healthcare and have a specimen collected for testing:

$$P(\text{Test}|\text{ILI}, week = w) \sim \text{Beta}(\alpha_w, \beta_w) \tag{6}$$

$$\mathbb{E}\left[\text{Test}|\text{ILI}, week = w\right] = \frac{\alpha_{w-1} + T_w}{\alpha_{w-1} + \beta_{w-1} + N_w} = \frac{\alpha_w}{\alpha_w + \beta_w} \tag{7}$$

$$\alpha_w = \alpha_{w-1} + T_w \tag{8}$$

$$\beta_w = \beta_{w-1} + N_w - T_w \tag{9}$$

Note that as the total number of reported ILI cases increases, the expected value of the posterior approaches the observed proportion of ILI cases who reported having a specimen collected:

$$\lim_{w\to\infty} \mathbb{E}\left[\text{Test}|\text{ILI}\right] = \lim_{w\to\infty} \frac{\alpha_0 + \sum_{i=1}^{w} T_w}{\alpha_0 + \beta_0 + \sum_{i=i}^{w} N_w} = \frac{T}{N} \tag{10}$$

4.3. Influenza surveillance data

For three of the cities for which we generated forecasts — Brisbane, Melbourne, and Sydney — we obtained influenza case notifications data for the 2014–2017 calendar years from the relevant data custodian. These data include only those cases which meet the Communicable Diseases Network Australia (CDNA) case definition for laboratory-confirmed influenza [37], and represent only a small proportion of the actual cases in the community [38].

For Perth, we obtained PCR-confirmed diagnoses from PathWest Laboratory Medicine WA (the Western Australia public pathology laboratory) for the 2014–2017 calendar years. In comparison to the notifications data obtained for the other cities in this study, these data do not include serologic diagnoses. They also contain a much higher proportion of hospitalised cases (40–60%), which is further magnified because the bulk of PathWest community specimens come from regional and remote areas. People whose symptoms are sufficiently severe to result in hospitalisation will most likely seek healthcare, regardless of any other behavioural influences, and so these data are likely to exhibit little sensitivity to behaviour changes characterised in the Flutracking survey data.

Many specimen do not include a symptom onset date, in which case we used the specimen collection date as a proxy. The difference between the time of symptom onset and the collection date is assumed to be less than one week for the majority of cases, and since we aggregate the case counts by week, we assume that this does not substantially impact the resulting time-series data.

Note that in this study we used the weekly case counts as reported after the 2017 influenza season had finished and all tests had been processed.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4.4. Forecasting methods

We generated near-real-time forecasts by combining an $SEIR$ compartment model of infection with influenza case counts through the use of a bootstrap particle filter [3,9,23].

First, we constructed a multivariate normal distribution for the peak timing and size in years prior to 2017, to characterise our prior expectation for the 2017 season. Parameter values for the $SEIR$ model were then sampled from uniform distributions, and each sample was accepted in proportion to the simulated epidemic's probability density. See Appendix B for details.

We modelled the relationship between disease incidence in the $SEIR$ model and the influenza case counts using a negative binomial distribution with dispersion parameter k, since the data are non-negative integer counts and are over-dispersed when compared to a Poisson distribution [3].

The probability of being *observed* (i.e., of being reported as a notifiable case) was the product of two probabilities: that of becoming infectious (p_{inf}), and that of being identified ($p_{\text{id}}(w)$, the probability of being symptomatic, presenting to a doctor, and having a specimen collected). The probability of becoming infectious was defined as the fraction of the *model* population that became infectious (i.e., transitioned from E to I), and subsumed symptomatic and asymptomatic infections:

$$p_{\text{inf}}(w) = S(w-1) + E(w-1) - S(w) - E(w) \tag{11}$$

$$p_{\text{ili}}(w) = p_{\text{inf}}(w) \cdot p_{\text{id}}(w) + [1 - p_{\text{inf}}(w)] \cdot p_{\text{bg}} \tag{12}$$

Values for $p_{\text{id}}(0)$ and k were informed by retrospective forecasts using surveillance data from previous seasons [3], while the background rate p_{bg} was estimated from out-of-season levels (March to May, 2017). The climatic modulation signals $F(w)$ were characterised by smoothed absolute humidity data for each city in previous years, as previously described [24].

To calibrate the forecasts at each week w, we scaled the reference testing probability $p_{\text{id}}(0)$ in proportion to the change in the expected testing probability (relative to the prior expectation), using a scaling coefficient κ. When the change was negative (i.e., the expected testing probability was less than the prior expectation) we instead raised the change to the power κ, to ensure that the calibrated value $p_{\text{id}}(w)$ was strictly positive:

$$p_{\text{id}}(w) = p_{\text{id}}(0) \times \left[1 + f_\kappa\left(\frac{\mathbb{E}\left[\text{Test}|\text{ILI}, week = w\right] - \mu_0}{\mu_0}\right)\right] \tag{13}$$

$$f_\kappa(x) = \begin{cases} \kappa \cdot x & \text{if } x \geq 0 \\ (x+1)^\kappa - 1 & \text{otherwise} \end{cases} \tag{14}$$

As required, this is a smooth function of the change in the expected testing probability. We recover the non-calibrated forecasts by setting $\kappa = 0$, in which case $p_{\text{id}}(w) = p_{\text{id}}(0)$, as shown in Figure 3.

4.5. Forecast performance

Forecast performance for each value of κ was measured via Bayes factors [24,25], relative to the non-calibrated forecasts. This is defined as the ratio $\widetilde{B}_k(w)$, for each week w, of (a) the likelihood of the future case counts $y_{w+1:N}$ according to the calibrated forecasts ($\kappa = k$); and (b) the likelihood of the future case counts according to the non-calibrated forecasts ($\kappa = 0$):

$$\widetilde{B}_k(w) = \frac{\mathbb{P}(y_{w+1:N}|y_{0:w}, \kappa = k)}{\mathbb{P}(y_{w+1:N}|y_{0:w}, \kappa = 0)} \tag{15}$$

More generally, we can measure forecast performance relative to any value K for κ:

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\widetilde{B}_{k,K}(w) = \frac{\mathbb{P}(y_{w+1:N}|y_{0:w}, \kappa = k)}{\mathbb{P}(y_{w+1:N}|y_{0:w}, \kappa = K)} \tag{16}$$

Our interest is in the average performance over the set of all forecasting weeks W:

$$\widetilde{B}_k = \frac{1}{|W|} \sum_{w \in W} \widetilde{B}_k(w) \tag{17}$$

$$\widetilde{B}_{k,K} = \frac{1}{|W|} \sum_{w \in W} \widetilde{B}_{k,K}(w) \tag{18}$$

Note that the forecasting weeks W are only a subset of those weeks for which we have Flutracking survey data and influenza surveillance data.

Values in excess of 10 are considered as positive or strong evidence that the chosen model is more strongly supported by the data, and values in excess of 100 are considered as very strong ("decisive") evidence [25,39]. Likewise, values less than -10 and less than -100 indicate strong and very strong evidence that the non-calibrated forecasts are more strongly supported by the data.

Here we consider the logarithms to base 10 of $\widetilde{B}_k$ and $\widetilde{B}_{k,K}$, which we respectively denote $\widetilde{b}_k$ and $\widetilde{b}_{k,K}$. The evidence thresholds for $\widetilde{b}_k$ (and also $\widetilde{b}_{k,K}$) are therefore:

$$|\widetilde{b}_k| \geq 1 \implies \text{strong evidence} \tag{19}$$

$$|\widetilde{b}_k| \geq 2 \implies \text{very strong evidence} \tag{20}$$

where the sign indicates whether the data support the calibrated forecasts (positive values) or the non-calibrated forecasts (negative values).

To measure the *relative* performance of different values for κ, we also define the *normalised* logarithms $\widehat{b}_k$ and $\widehat{b}_{k,K}$, expressed as percentages of the maximum absolute value obtained over the chosen set of values V for κ:

$$\widehat{b}_k = 100 \times \frac{\widetilde{b}_k}{|\widetilde{b}_M|} \qquad M = \arg\max_{v \in V} |\widetilde{b}_v| \tag{21}$$

$$\widehat{b}_{k,K} = 100 \times \frac{\widetilde{b}_{k,K}}{|\widetilde{b}_{N,K}|} \qquad N = \arg\max_{v \in V} |\widetilde{b}_{v,K}| \tag{22}$$

4.6. Availability of materials and methods

The data and source code used to perform all of the forecast simulations and generate the results presented in this manuscript are published online at doi:10.26188/5bf3d60c4ffae, under the terms of permissive licenses.

Author Contributions: Conceptualization, R.M., A.E.Z., S.J.C., and J.M.M.; Methodology, R.M., A.E.Z., S.J.C, and J.M.M; Software, R.M.; Validation, R.M.; Formal Analysis, R.M.; Investigation, R.M., A.E.Z., S.J.C., and J.M.M.; Resources, S.J.C. and J.M.M.; Data Curation, R.M. and S.J.C.; Writing – Original Draft Preparation, R.M.; Writing – Review & Editing, R.M., A.E.Z., S.J.C., and J.M.M.; Visualization, R.M., A.E.Z., S.J.C., and J.M.M.; Supervision, R.M.; Project Administration, R.M; Funding Acquisition, R.M. and J.M.M.

Funding: This research was funded by the Defence Science and Technology Group project "Bioterrorism Preparedness Strategic Research Initiative 07/301", and by the Australian National Health and Medical Research Council (NHMRC) Centre for Research Excellence (CRE) grant 1078068.

Acknowledgments: We are extremely grateful to Tony Lau and Peter Dawson (Land Personnel Protection Branch, Land Division, Defence Science and Technology Group) for their ongoing support, assistance, and involvement in this work.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We would like to acknowledge the active and ongoing contributions from public health staff in the participating jurisdictions, without whom this work would not have been possible:

- Robin Gilmour and Nathan Saul. Communicable Disease Branch, Health Protection New South Wales, New South Wales.
- Frances A. Birrell, Angela Wakefield, and Mayet Jayloni. Epidemiology and Research Unit, Communicable Diseases Branch, Prevention Division, Department of Health, Queensland.
- Lucinda J. Franklin, Nicola Stephens, Janet Strachan, Trevor Lauer, and Kim White. Communicable Diseases Section, Health Protection Branch, Regulation Health Protection and Emergency Management Division, Victorian Government Department of Health and Human Services, Victoria.
- Avram Levy and Cara Minney-Smith. PathWest Laboratory Medicine WA, Department of Health, Western Australia.

We would also like to acknowledge the Flutracking team for their participation and support, and for organising the "Working with Flutracking Data" workshop held on 2017-06-29. Finally, we would like to acknowledge all of the participants who complete the weekly Flutracking surveys and provide us with this valuable data set.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CDNA	Communicable Diseases Network Australia
ILI	Influenza-like illness
RT-PCR	Reverse transcription polymerase chain reaction
SEIR	Susceptible, exposed, infectious, recovered

Appendix A. Flutracking statistics

Table A1. Flutracking mean weekly count of completed surveys, participants with ILI and participants with ILI who sought health advice and reported being tested for influenza, 2014-2017, NSW, Vic., Qld, and WA.

State	Year	Completed surveys	Participants with ILI	Reported tests
NSW	2014	6311	146	3
	2015	7242	155	4
	2016	8439	178	5
	2017	9405	218	14
Vic	2014	2815	64	1
	2015	3348	68	2
	2016	3894	75	1
	2017	4712	104	5
Qld	2014	1736	36	1
	2015	2180	46	2
	2016	2481	50	2
	2017	2822	68	4
WA	2014	1039	24	1
	2015	3804	88	1
	2016	3511	90	2
	2017	3541	72	2

Appendix B. Forecasting methods

We applied the forecasting method used in our previous forecasting studies [3,9,10,23,24], which combines an $SEIR$ compartment model of infection with influenza case counts through the use of a bootstrap particle filter.

In brief, we randomly selected parameter values (Table A2) for the $SEIR$ model (Eqs A1–A6) to generate a suite of candidate epidemics, and defined a relationship between disease incidence in the

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SEIR model and the influenza case counts (an "observation model", Eqs A7–A11). This allowed us to calculate the likelihood of "observing" the case counts from each of the model simulations. The particle filter (Eqs A12–A16) was used to update these likelihoods ("weights" w_i) as data ($\mathbf{y_t}$) were reported, and to discard extremely unlikely simulations in favour of more likely ones ("resampling").

For each city, we constructed a multivariate normal distribution from the observed peak timing and size in each of the 2012–16 influenza seasons (Table A3). For each set of sampled parameter values, we simulated the epidemic that those values described and accepted the sample in proportion to the simulated epidemic's probability density according to this multivariate distribution [10].

$$\frac{dS}{dt} = -\beta SI - \theta_{seed} \tag{A1}$$

$$\frac{dE}{dt} = \beta SI + \theta_{seed} - \sigma E \tag{A2}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{A3}$$

$$\frac{dR}{dt} = \gamma I \tag{A4}$$

$$\beta = R_0 \cdot \gamma \cdot [1 + \alpha \cdot F(t)] \tag{A5}$$

$$\theta_{\text{seed}} = \begin{cases} \frac{1}{N} & \text{if } S(t) = 1 \text{ and } t = \tau \\ 0 & \text{otherwise} \end{cases} \tag{A6}$$

$$P(\mathbf{y_t} \mid \mathbf{x_t}; k) = \frac{\Gamma(\mathbf{y_t} + k)}{\Gamma(k) \cdot \mathbf{y_t}!} \cdot (p_{\text{k}})^k \cdot (1 - p_{\text{k}})^{\mathbf{y_t}} \tag{A7}$$

$$p_{\text{k}} = \frac{k}{k + N \cdot p_{\text{ili}}} \tag{A8}$$

$$p_{\text{ili}}(t) = p_{\text{inf}}(t) \cdot p_{\text{id}}(t) + [1 - p_{\text{inf}}(t)] \cdot p_{\text{bg}} \tag{A9}$$

$$p_{\text{bg}} = bg \cdot N^{-1} \tag{A10}$$

$$p_{\text{inf}}(t) = S(t - \Delta) + E(t - \Delta) - S(t) - E(t) \tag{A11}$$

$$\mathbf{x_t} = [S(t), E(t), I(t), R(t), R_0, \alpha, \sigma, \gamma, \tau]^T \tag{A12}$$

$$w_i(0) = \left(N_{\text{px}}\right)^{-1} \tag{A13}$$

$$w_i'(t \mid \mathbf{y_t}) = w_i(t-1) \cdot P(\mathbf{y_t} \mid \mathbf{x_t^i}; k) \tag{A14}$$

$$w_i(t \mid \mathbf{y_t}) = w_i'(t) \cdot \left(\sum_{j=1}^{N_{\text{px}}} w_j'(t) \right)^{-1} \tag{A15}$$

$$N_{\text{eff}}(t) = \left(\sum_{j=1}^{N_{\text{px}}} \left[w_j'(t) \right]^2 \right)^{-1} \tag{A16}$$

As in our previous studies, we assumed the relationship between disease incidence in the *SEIR* model and the weekly case counts was defined by a negative binomial distribution with dispersion parameter k (Eqs A7–A8), since the data are non-negative integer counts and are over-dispersed when compared to a Poisson distribution. The probability of being *observed* (i.e., of being reported as a notifiable case, Eq A9) was the product of two probabilities: that of becoming infectious (p_{inf}), and that of being identified (p_{id}, the likelihood of being symptomatic, presenting to a doctor, and having a specimen collected). The probability of becoming infectious was defined as the fraction of the *model* population that became infectious (i.e., transitioned from E to I), and subsumed symptomatic and asymptomatic infections (Eq A11).

Values for $p_{\text{id}}(0)$ and k were informed by retrospective forecasts using surveillance data from previous seasons [3], while the background rate p_{bg} was estimated from out-of-season levels (March

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to May) in each year. The climatic modulation signals $F(t)$ were characterised by smoothed absolute humidity data for each city in previous years, as previously described [24].

Forecasts were generated using the `pypfilt` and `epifx` packages for Python, which were developed as part of this project and are distributed under permissive free software licenses. Installation instructions and tutorial documentation are available at https://epifx.readthedocs.io/.

Table A2. Parameter values for (i) the transmission model; (ii) the bootstrap particle filter; and (iii) the observation model. Here, $\mathcal{U}(x, y)$ denotes the continuous uniform distribution and $\mathcal{U}\{x, y\}$ denotes the discrete uniform distribution, where x and y are the minimum and maximum values, respectively.

	Meaning	Value
β	Force of infection	Eq A5
R_0	Basic reproduction number	$\sim \mathcal{U}(1.0, 1.6)$
σ	Inverse of the incubation period (days^{-1})	$\sim [\mathcal{U}(0.33, 2.0)]^{-1}$
γ	Inverse of the infectious period (days^{-1})	$\sim [\mathcal{U}(0.33, 1.0)]^{-1}$
α	Seasonal modulation of R_0	$\sim \mathcal{U}(-0.2, 0.0)$
τ	Time of the first infection (days)	$\sim \mathcal{U}\{0, 50\}$
N_{px}	Number of particles (simulations)	$15,000$
N_{min}	Minimum number of effective particles	$0.25 \cdot N_{\text{px}}$
Δ	Observation period (days)	7
k	Dispersion parameter	100
p_{bg}	Background observation probability	Eq A10
bg	Estimated background observation rate	see Table A3
p_{id}	Observation probability	see Table A3
N	Population size	see Table A3

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table A3. The forecast parameter values, and the observed peak timing and size, for each of the capital cities. The observed peaks for 2012–16 were used to construct a multivariate normal distribution for the peak timing and size in each city, in order to characterise the model priors for the 2017 season.

City	Year	Peak size	Peak timing	bg	$p_{\mathrm{id}}(0)$	N
Brisbane	2012	510	Aug 12			
	2013	140	Sep 1			
	2014	509	Aug 24	28	0.003	2,308,700
	2015	986	Aug 23	48	0.003	2,308,700
	2016	544	Sep 4	46	0.003	2,308,700
	2017	1346	Aug 27	40	0.003	2,308,700
Melbourne	2012	360	Jul 15			
	2013	417	Aug 25			
	2014	691	Aug 24	36	0.00275	4,108,541
	2015	1329	Aug 30	73	0.00275	4,108,541
	2016	975	Sep 4	62	0.00275	4,108,541
	2017	3962	Sep 10	79	0.00275	4,108,541
Perth	2012	351	Jul 15			
	2013	106	Sep 15			
	2014	199	Aug 24	10	0.002	1,900,000
	2015	170	Aug 23	20	0.002	1,900,000
	2016	221	Aug 14	14	0.002	1,900,000
	2017	94	Sep 10	5	0.002	1,900,000
Sydney	2012	535	Jul 1			
	2013	855	Sep 1			
	2014	1989	Aug 17	38	0.0076	4,921,000
	2015	2933	Aug 23	75	0.0076	4,921,000
	2016	2884	Sep 4	138	0.0076	4,921,000
	2017	8798	Aug 20	139	0.0076	4,921,000

1. Reed, C.; Chaves, S.S.; Daily Kirley, P.; Emerson, R.; Aragon, D.; Hancock, E.B.; Butler, L.; Baumbach, J.; Hollick, G.; Bennett, N.M.; et al.. Estimating Influenza Disease Burden from Population-Based Surveillance Data in the United States. *PLOS ONE* **2015**, *10*, e0118369. doi:10.1371/journal.pone.0118369.
2. Biggerstaff, M.; Alper, D.; Dredze, M.; Fox, S.; Fung, I.C.H.; Hickmann, K.S.; Lewis, B.; Rosenfeld, R.; Shaman, J.; Tsou, M.H.; others. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infectious Diseases* **2016**, *16*, 357. doi:10.1186/s12879-016-1669-x.
3. Moss, R.; Zarebski, A.; Dawson, P.; McCaw, J.M. Retrospective forecasting of the 2010–14 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology and Infection* **2017**, *145*, 156–169. doi:10.1017/S0950268816002053.
4. Riley, P.; Ben-Nun, M.; Turtle, J.A.; Linker, J.; Bacon, D.P.; Riley, S. Identifying factors that may improve mechanistic forecasting models for influenza. *bioRxiv* **2017**, p. 172817. doi:10.1101/172817.
5. Ben-Nun, M.; Riley, P.; Turtle, J.; Bacon, D.; Riley, S. National and Regional Influenza-Like-Illness Forecasts for the USA. *bioRxiv* **2018**, p. 309021.
6. Biggerstaff, M.; Johansson, M.; Alper, D.; Brooks, L.C.; Chakraborty, P.; Farrow, D.C.; Hyun, S.; Kandula, S.; McGowan, C.; Ramakrishnan, N.; Rosenfeld, R.; Shaman, J.; Tibshirani, R.; Tibshirani, R.J.; Vespignani, A.; Yang, W.; Zhang, Q.; Reed, C. Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics* **2018**. doi:10.1016/j.epidem.2018.02.003.
7. Cope, R.C.; Ross, J.V.; Chilver, M.; Stocks, N.P.; Mitchell, L. Characterising seasonal influenza epidemiology using primary care surveillance data. *PLOS Computational Biology* **2018**, *14*, e1006377. doi:10.1371/journal.pcbi.1006377.
8. Cope, R.C.; Ross, J.V.; Chilver, M.; Stocks, N.P.; Mitchell, L. Connecting surveillance and population-level influenza incidence. *bioRxiv* **2018**, p. 427708.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

9. Moss, R.; Fielding, J.E.; Franklin, L.J.; Stephens, N.; McVernon, J.; Dawson, P.; McCaw., J.M. Epidemic forecasts as a tool for public health: interpretation and (re)calibration. *Australian and New Zealand Journal of Public Health* **2018**, *42*, 69–76. doi:10.1111/1753-6405.12750.
10. Moss, R.; Zarebski, A.E.; Dawson, P.; Franklin, L.J.; Birrell, F.A.; McCaw., J.M. Anatomy of a seasonal influenza epidemic forecast. *Communicable Diseases Intelligence* **2018**. Under review.
11. Funk, S.; Bansal, S.; Bauch, C.T.; Eames, K.T.; Edmunds, W.J.; Galvani, A.P.; Klepac, P. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics* **2015**, *10*, 21–25. doi:10.1016/j.epidem.2014.09.005.
12. Moran, K.R.; Fairchild, G.; Generous, N.; Hickmann, K.; Osthus, D.; Priedhorsky, R.; Hyman, J.; Del Valle, S.Y. Epidemic Forecasting is Messier Than Weather Forecasting: The Role of Human Behavior and Internet Data Streams in Epidemic Forecast. *Journal of Infectious Diseases* **2016**, *214*, S404–S408. doi:10.1093/infdis/jiw375.
13. Mitchell, L.; Ross, J.V. A data-driven model for influenza transmission incorporating media effects. *Royal Society Open Science* **2016**, *3*, 160481. doi:10.1098/rsos.160481.
14. Reich, N.G.; Brooks, L.; Fox, S.; Kandula, S.; McGowan, C.; Moore, E.; Osthus, D.; Ray, E.L.; Tushar, A.; Yamana, T.; others. Forecasting seasonal influenza in the US: A collaborative multi-year, multi-model assessment of forecast performance. *bioRxiv* **2018**, p. 397190. doi:10.1101/397190.
15. Kandula, S.; Yamana, T.; Pei, S.; Yang, W.; Morita, H.; Shaman, J. Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *Journal of the Royal Society, Interface* **2018**, *15*. doi:10.1098/rsif.2018.0174.
16. Moss, R.; McCaw, J.M.; Cheng, A.C.; Hurt, A.C.; McVernon, J. Reducing disease burden in an influenza pandemic by targeted delivery of neuraminidase inhibitors: mathematical models in the Australian context. *BMC Infectious Diseases* **2016**, *16*. doi:10.1186/s12879-016-1866-7.
17. Carlson, S.J.; Dalton, C.B.; Durrheim, D.N.; Fejsa, J. Online Flutracking Survey of Influenza-like Illness during Pandemic (H1N1) 2009, Australia. *Emerging Infectious Diseases* **2010**, *16*, 1960–1962. doi:10.3201/eid1612.100935.
18. Carlson, S.J.; Dalton, C.B.; Butler, M.T.; Fejsa, J.; Elvidge, E.; Durrheim, D.N. Flutracking weekly online community survey of influenza-like illness annual report 2011 and 2012. *Communicable Diseases Intelligence* **2013**, *37*, E398–E406.
19. Dalton, C.B.; Carlson, S.J.; McCallum, L.; Butler, M.T.; Fejsa, J.; Elvidge, E.; Durrheim, D.N. Flutracking weekly online community survey of influenza-like illness: 2013 and 2014. *Communicable Diseases Intelligence* **2015**, *39*, E361–E368.
20. Dalton, C.B.; Carlson, S.J.; Durrheim, D.N.; Butler, M.T.; Cheng, A.C.; Kelly, H.A. Flutracking weekly online community survey of influenza-like illness annual report, 2015. *Communicable Diseases Intelligence* **2016**, *40*, E512–E520.
21. Department of Health. Australian Influenza Surveillance Report: 2017 Season Summary. Technical report, Australian Government, 2017.
22. Scutti, S. 'Australian flu': It's not from Australia. *CNN* **2018**. [accessed on 2018-10-05: https://edition.cnn.com/2018/02/02/health/australian-flu-became-global/index.html].
23. Moss, R.; Zarebski, A.; Dawson, P.; McCaw, J.M. Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza and Other Respiratory Viruses* **2016**, *10*, 314–323. doi:10.1111/irv.12376.
24. Zarebski, A.E.; Dawson, P.; McCaw, J.M.; Moss, R. Model selection for seasonal influenza forecasting. *Infectious Disease Modelling* **2017**, *2*, 56–70. doi:10.1016/j.idm.2016.12.004.
25. Kass, R.E.; Raftery, A.E. Bayes factors. *Journal of the American Statistical Association* **1995**, *90*, 773–795. doi:10.1080/01621459.1995.10476572.
26. Ibuka, Y.; Chapman, G.B.; Meyers, L.A.; Li, M.; Galvani, A.P. The dynamics of risk perceptions and precautionary behavior in response to 2009 (H1N1) pandemic influenza. *BMC Infectious Diseases* **2010**, *10*, 296. doi:10.1186/1471-2334-10-296.
27. Davis, M.D.M.; Stephenson, N.; Lohm, D.; Waller, E.; Flowers, P. Beyond resistance: social factors in the general public response to pandemic influenza. *BMC Public Health* **2015**, *15*, 436. doi:10.1186/s12889-015-1756-8.

bioRxiv preprint doi: https://doi.org/10.1101/475756; this version posted November 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

28. De Angelis, D.; Presanis, A.M.; Birrell, P.J.; Tomba, G.S.; House, T. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics* **2015**, *10*, 83–87. doi:10.1016/j.epidem.2014.09.004.
29. Presanis, A.M.; Pebody, R.G.; Paterson, B.J.; Tom, B.D.M.; Birrell, P.J.; Charlett, A.; Lipsitch, M.; De Angelis, D. Changes in severity of 2009 pandemic A/H1N1 influenza in England: a Bayesian evidence synthesis. *BMJ* **2011**, *343*, d5408. doi:10.1136/bmj.d5408.
30. Birrell, P.J.; De Angelis, D.; Presanis, A.M. Evidence synthesis for stochastic epidemic models. *Statistical Science* **2018**, *33*, 34–43. doi:10.1214/17-STS631.
31. van Noort, S.P.; Codeço, C.T.; Koppeschaar, C.E.; van Ranst, M.; Paolotti, D.; Gomes, M.G.M. Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. *Epidemics* **2015**, *13*, 28–36. doi:10.1016/j.epidem.2015.05.001.
32. Baltrusaitis, K.; Santillana, M.; Crawley, A.W.; Chunara, R.; Smolinski, M.; Brownstein, J.S. Determinants of Participants' Follow-Up and Characterization of Representativeness in Flu Near You, A Participatory Disease Surveillance System. *JMIR Public Health and Surveillance* **2017**, *3*, e18. doi:10.2196/publichealth.7304.
33. Chretien, J.P.; George, D.; Shaman, J.; Chitale, R.A.; McKenzie, F.E. Influenza forecasting in human populations: a scoping review. *PLOS ONE* **2014**, *9*, e94130. doi:10.1371/journal.pone.0094130.
34. Nsoesie, E.O.; Brownstein, J.S.; Ramakrishnan, N.; Marathe, M.V. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses* **2014**, *8*, 309–316. doi:10.1111/irv.12226.
35. Priedhorsky, R.; Osthus, D.; Daughton, A.R.; Moran, K.R.; Culotta, A. Deceptiveness of internet data for disease surveillance. *arXiv e-prints* **2017**, [1711.06241v1]. doi:https://arxiv.org/abs/1711.06241v1.
36. Domnich, A.; Panatto, D.; Signori, A.; Lai, P.L.; Gasparini, R.; Amicizia, D. Age-Related Differences in the Accuracy of Web Query-Based Predictions of Influenza-Like Illness. *PLOS ONE* **2015**, *10*, e0127754. doi:10.1371/journal.pone.0127754.
37. Rowe, S.L. Infectious diseases notification trends and practices in Victoria, 2011. *Victorian Infectious Diseases Bulletin* **2012**, *15*, 92–97.
38. Dalton, C.B.; Carlson, S.J.; Butler, M.T.; Elvidge, E.; Durrheim, D.N. Building Influenza Surveillance Pyramids in Near Real Time, Australia. *Emerging Infectious Diseases* **2013**, *19*, 1863–1865. doi:10.3201/eid1911.121878.
39. Jeffreys, H. *The theory of probability*; OUP Oxford, 1998.