Forecasting influenza in Europe and globally: the role of absolute humidity and human travel, and the potential for use in public health decision making

Sarah Corinne Kramer

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

Forecasting influenza in Europe and globally: the role of absolute humidity and human travel, and the potential for use in public health decision making

Sarah Corinne Kramer

Influenza causes substantial morbidity and mortality yearly in both temperate and tropical regions, as well as sporadic and potentially severe pandemics. Although vaccines for seasonal influenza exist, most options for controlling influenza outbreaks are reactive in nature. Sufficiently accurate and well-calibrated forecasts, on the other hand, could allow public health practitioners, medical professionals, and the public to respond to unfolding influenza outbreaks proactively. For example, hospitals could prepare additional beds for a predicted surge, and public health experts could redouble vaccination efforts. Recently, skillful forecasts have been developed for a range of infectious diseases, including influenza, but this work has been limited to only a few countries. In this dissertation, we explore the potential for generating accurate influenza forecasts using a publicly-available dataset of country-level epidemiologic and virologic surveillance data. In Chapter 2, we use a combined model-inference system to generate retrospective forecasts for 64 countries in both temperate and tropical climates. We show that forecast accuracy is significantly better in countries with temperate climates, and that inclusion of environmental forcing, specifically modulation of viral transmissibility due to variability of absolute humidity conditions, also improves forecast accuracy in temperate climates. In Chapter

3, we develop a metapopulation model of twelve European countries using data on international air travel and commuting. We find that this model is unable to produce more skillful forecasts than those produced for individual countries in isolation. We make recommendations for improvements in data collection and reporting that may increase the success of similar modeling efforts in the future. In Chapter 4, we assess the performance of real-time forecasts generated for 37 countries over two influenza seasons and discuss the potential for their use in public health decision making. Finally, in Chapter 5 we describe the results of a small survey of public health practitioners in the United States. We find that the majority of respondents desire more effective communication between modelers and public health practitioners, and we discuss the importance of regular and improved communication in advancing the practical use of forecasts as public health decision making tools. This dissertation advances the science of influenza forecasting by demonstrating that skillful retrospective and real-time forecasts can be generated for many countries where previous forecasting efforts are either minimal or absent. However, it is vital that data quality issues be addressed if further progress is to be made. Future work should focus in particular on climatic drivers of influenza in the tropics and subtropics, on the role of human travel at various spatial scales, and on the development of regional and local forecasting capacity. Additionally, dedicated collaboration between modelers and public health practitioners will be instrumental for motivating and informing the use of forecasts in combating influenza outbreaks.

Table of Contents

List of Tables and Figures

Tables:

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Figures:

Chapter 1

Chapter 2

Chapter 4

Chapter 5

Chapter 6

List of Abbreviations

A(H1)	Influenza A Virus Subtype H1N1
A(H3)	Influenza A Virus Subtype H3N2
aOR	Adjusted Odds Ratio
ARI	Acute Respiratory Infection
ASTHO	Association of State and Territorial Health Officials
CDC	Centers for Disease Control and Prevention
COVID-19	Coronavirus Disease 2019
CSTE	Council of State and Territorial Epidemiologists
EAKF	Ensemble Adjustment Kalman Filter
ECDC	European Centre for Disease Prevention and Control
EPI	Epidemic Prediction Initiative
GEE	Generalized Estimating Equation
GFT	Google Flu Trends
GISRS	Global Influenza Surveillance and Response System
GLDAS	Global Land Data Assimilation System
HA	Hemagglutinin
IIAG	Influenza Incidence Analytics Group
ILI	Influenza-Like Illness
MAE	Mean Absolute Error
NA	Neuraminidase
NACCHO	National Association of County and City Health Officials
OEV	Observational Error Variance
R_e/R_{eff}	Effective Reproductive Number
RMSE	Root Mean Square Error
R_0	Basic Reproductive Number
RSV	Respiratory Syncytial Virus
SARI	Severe Acute Respiratory Infection
SEIRS	Susceptible-Exposed-Infected-Recovered-Susceptible
SIRS	Susceptible-Infected-Recovered-Susceptible
(s)MAPE	(Symmetric) Mean Absolute Percentage Error
WHO	World Health Organization

Country Abbreviations

AT	Austria	HU	Hungary
BE	Belgium	IT	Italy
CZ	Czechia	LU	Luxembourg
DE	Germany	NL	Netherlands
ES	Spain	PL	Poland
FR	France	SK	Slovakia

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Jeffrey Shaman, for all of his guidance and encouragement over the past five years. It has been an honor to work with someone who is not only a talented scientist but also a compassionate mentor throughout my time at Columbia. Jeff, thank you for keeping me on track while giving me the freedom to explore my scientific interests.

I also would like to thank my thesis committee, Drs. Wan Yang, Simon Anthony, Micaela Martinez, and Marianthi Kioumourtzoglou for all of their suggestions and support as I worked on this dissertation, especially as I debated the seemingly infinite ways of proceeding with the metapopulation model. Wan and Simon, thank you for the opportunity to work on projects with you outside of my dissertation work. It has been so exciting to expand my knowledge and skills in this way. Dr. Greg Freyer, thank you so much for joining my committee at the last minute.

Dr. Sen Pei, I cannot thank you enough for all of your assistance as I constructed and worked with the European metapopulation model. I could not have accomplished this without your guidance, and I am so grateful that you were always willing to help.

I would also like to thank all the members of the Shaman research group, past and present. I am grateful to have spent the last five years getting to know you as scientists and as people. In particular, thank you to Dr. Alex Heaney for being a great mentor and friend, and for helping me through each and every step of this process. Sasi Kandula, thank you for all of your help with coding, and with getting my real-time forecasts online. Dr. Katherine Crocker, thank you for your mentorship as I finished my research and began writing my dissertation. Your

advice and understanding have been so important to me. Haruka Morita, thank you for answering my endless stream of questions over the past five years.

I'm grateful to my wonderful EHS cohort, Mike He, Roheeni Saxena, and Yanelli Núñez, for their support and friendship throughout this process. I feel so lucky to have worked toward this goal alongside such a talented and kindhearted group of people. Additionally, I'm grateful to Sebastian Rowland, Lizzy Gibson, and Nicole Comfort, both for their scientific input as well as for their continual friendship and encouragement. I could not have done this without the support of all of my EHS friends.

Nina Kulacki and Bernice Ramos-Perez, thank you for your constant assistance and unwavering patience in handling all of my departmental- and finance-related concerns over the years.

Thank you so much to my family for always believing in me. I wouldn't be where I am today without you encouraging me to pursue my goals throughout my life. I love you so much.

Last but not least, thank you to my husband, Ralf Strobel, for being there for me through all of the ups and downs of the past five years. You make me a better person, and I couldn't ask for a more supportive partner.

Chapter 1

Introduction

Influenza

Influenza is a respiratory disease of both annual and pandemic importance. Seasonal influenza, which occurs every year in most locations globally, leads to an estimated one billion cases and 290,000 to 650,000 deaths worldwide (Iuliano et al., 2018). Infection with influenza typically leads to symptoms including fever, cough, muscle pain, and fatigue (WHO, n.d.-b), although milder and even asymptomatic infections also occur (Galanti et al., 2019). While of minimal concern to young, healthy people, influenza may cause severe complications in very young children, those over the age of 50, people with chronic respiratory or cardiac conditions, and those who are immunosuppressed (Grohskopf et al., 2019; WHO, n.d.-b). These yearly epidemics also impose a substantial economic toll, on the order of 11 to 14 billion dollars each

year in the US alone (Carrat & Flahault, 2007; Putri et al., 2018). In addition to regular outbreaks, influenza is the cause of sporadic pandemics, including both the relatively mild 2009 "swine flu" pandemic (Dawood et al., 2012), and the disastrous 1918 "Spanish flu" pandemic, which caused an estimated 50 million deaths worldwide (Taubenberger & Morens, 2006).

While there is a myriad of existing control measures against influenza, the majority are reactive in nature. In other words, public health officials and medical practitioners must wait for a situation to escalate, then respond accordingly. In order to reduce morbidity and mortality more effectively, more proactive measures must be developed and effectively integrated into decision making. This dissertation focuses specifically on the emerging field of influenza forecasting.

Influenza virology

Influenza is caused by the influenza virus (Figure 1), a segmented RNA virus of the family Orthomyxoviridae (Payne, 2017). We focus here on the genera, or "types," A and B, which consistently circulate in human populations (Carrat & Flahault, 2007; Cox & Subbarao, 2000). The infectious potential of influenza viruses is largely driven by two surface proteins: hemagglutinin (HA), which binds to host cells and allows the virus entry; and neuraminidase (NA), which allows viruses to leave host cells after replication (Tamura et al., 2005; Wille & Holmes, 2019). These surface proteins show a high degree of diversity among influenza A viruses, and determine to which "subtype" the virus belongs. For example, H3N2 has HA belonging to subtype 3 and NA belonging to subtype 2 (Carrat & Flahault, 2007; Cox & Subbarao, 2000). To date, eighteen subtypes of HA and eleven of NA have been discovered, most of which circulate among wild water- and shorebirds, the natural reservoir of type A

influenza (CDC, n.d.-b). Currently, only subtypes H1N1 and H3N2 circulate regularly among humans.

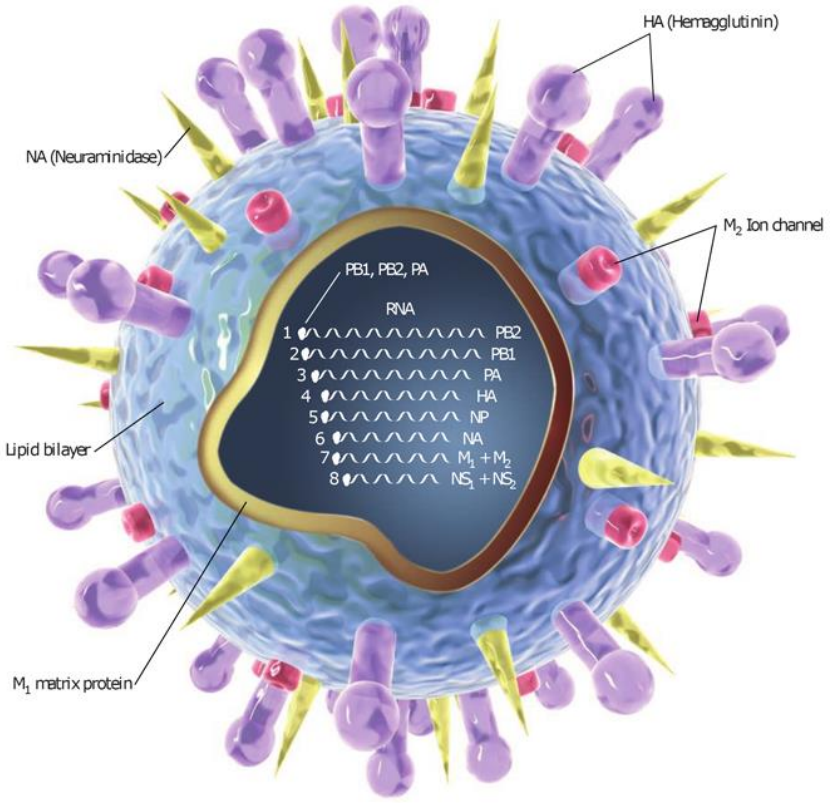


Figure 1. The influenza virus. Credit: C. Bickel/*Science*. From J. Kaiser. *Science* **312,** 380-2 (2006). Reprinted with permission from AAAS. NP: nucleoprotein; NS: non-structural protein; PA: polymerase A; PB1: polymerase B1; PB2: polymerase B2 (WHO, 2011a)

Influenza seasonality

In temperate regions, seasonal influenza outbreaks occur during the winter, with very few cases observed during the summer (Bloom-Feshbach et al., 2013). In the tropics and subtropics, influenza may cause biannual outbreaks, or else transmit continually year-round (Figure 2) (Bloom-Feshbach et al., 2013; WHO, 2011a). Notably, these differences in seasonality do not mean that the burden of influenza is lower in the tropics and subtropics (Ng & Gordon, 2015; L. Yang et al., 2011), and in fact mortality rates from seasonal influenza are highest in sub-Saharan Africa and southeast Asia (Iuliano et al., 2018). The exact drivers of these patterns are not known, and a range of hypotheses have been put forward (Tamerius et al., 2013). In temperate regions, absolute humidity, defined as the amount of water vapor in the air, appears to play a significant role, with lower humidity associated with higher rates of virus survival and transmission (Barreca & Shimshack, 2012;

Shaman & Kohn, 2009). Climatic drivers of outbreak patterns in the tropics and subtropics, where absolute humidity is consistently high, are less well-understood.

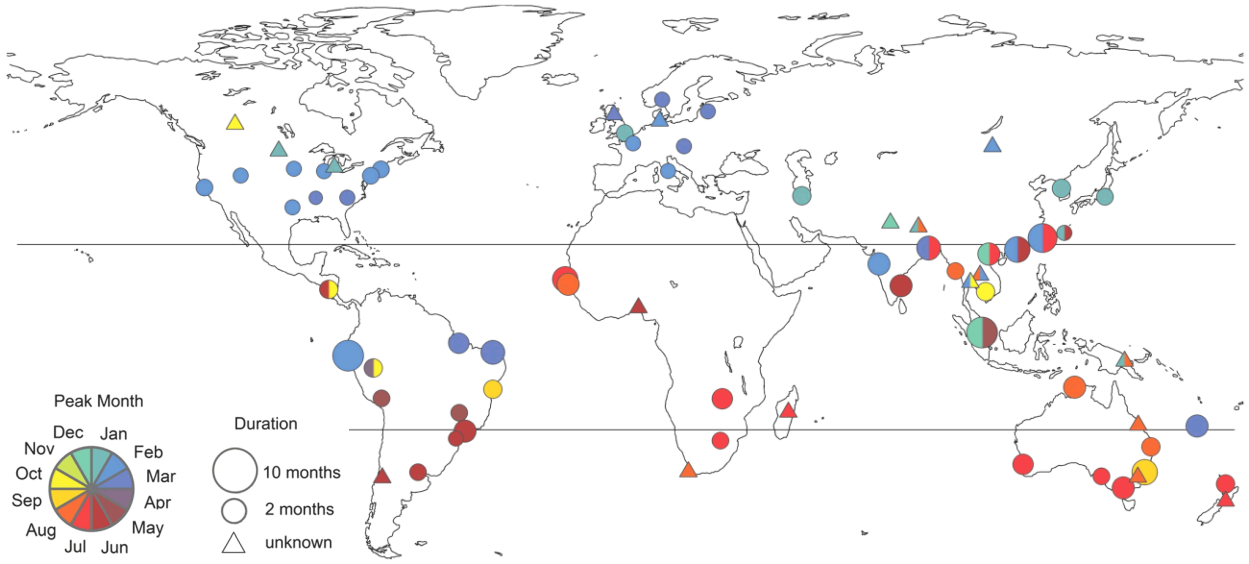


Figure 2. Seasonality of influenza outbreaks worldwide.
This figure was adapted from Bloom-Feshbach et al. (2013), Figure 2, under CC Public Domain license. It reflects their work determining the month of peak timing and the duration of influenza outbreaks in 77 locations. The color of the circles and triangles signifies the month of peak influenza activity, while circle size represents outbreak duration; for locations where several years of data were available, these values are averages. Points shaded with two colors indicate locations with two distinct peaks. The two horizontal black lines represent the boundaries of the tropics.

Seasonal influenza patterns are also driven to some extent by host susceptibility. While infection with influenza confers longstanding immunity to that particular strain (Carrat & Flahault, 2007), there is likely little cross-immunity between subtypes (Tamura et al., 2005), and rapid evolution leads influenza viruses to achieve significant antigenic change roughly every three to five years, on average (Liu et al., 2015; D. J. Smith, 2004). This immune escape, in conjunction with the simultaneous circulation of multiple (sub)types, allows regular outbreaks of influenza to occur (Carrat & Flahault, 2007; Cox & Subbarao, 2000), and leads to sometimes complicated (sub)type dynamics, with different (sub)types dominating depending on time and location (Yaari et al., 2013).

Despite the regularity with which influenza outbreaks occur in temperate regions, outbreak timing and intensity varies by season (Bresee & Hayden, 2013; CDC, n.d.-c). Preparing for influenza outbreaks in the tropics and subtropics, where there may be no strong pattern to outbreak timing, presents additional difficulties (Caini et al., 2016; Hirve et al., 2016).

Human travel and influenza transmission

Rapidly increasing global connectivity has greatly impacted the ways in which infectious diseases spread through human populations (Colizza et al., 2007). Still, the exact impact of different types of travel on disease transmission dynamics, as well as how these effects vary by location and spatial scale, are not well understood. Broadly speaking, long-distance air travel drives the introduction of pathogens into distant locations (Brockmann & Helbing, 2013; Brownstein et al., 2006; Charu et al., 2017; Lemey et al., 2014), and has been integral in the rapid global spread of both the 2009 influenza pandemic (Khan et al., 2009) and coronavirus disease 2019 (COVID-19) (De Salazar et al., 2020). Meanwhile, shorter-distance travel, such as automobile or train travel, are responsible for local transmission after introduction, as well as for seeding infection in locations that are not well-connected to air networks (Balcan et al., 2009; Charu et al., 2017). Commuting could, in theory, play a particularly important role in shorter-distance transmission because it represents regular travel, rather than an isolated event; commuters may therefore be more likely to be involved in transmission between locations. Studies in the US (Bozick & Real, 2015; Viboud et al., 2006) and France (Charaudeau et al., 2014) support the idea that commuting is an important driver of influenza transmission. However, other studies have found conflicting results (Crepey & Barthelemy, 2007), and these patterns do not necessarily hold for all countries (Geoghegan et al., 2018).

Potential for pandemic emergence

Although this dissertation focuses primarily on seasonal influenza, the potentially severe impact of future influenza pandemics warrants discussion. A total of four influenza pandemics have occurred since 1918 (Wille & Holmes, 2019), and the World Health Organization (WHO) considers the future emergence of a novel, pandemic influenza virus to be an absolute certainty (WHO, 2018a). Currently, while there are sporadic small outbreaks of avian influenza, none of the viruses have demonstrated the capacity for sustained human-to-human transmission (WHO, 2018b). However, evidence suggests only a small number of mutations could allow these viruses to spread between humans (Herfst et al., 2012; Imai et al., 2012). Furthermore, the segmented nature of the influenza genome allows for reassortment, or the exchange of entire genome segments, in an organism simultaneously infected with multiple influenza subtypes (Tamura et al., 2005; Wille & Holmes, 2019). If the HA or NA segments of a (sub)type with little or no existing human immunity are transferred to an existing human influenza strain, pandemic emergence can occur (Carrat & Flahault, 2007). Indeed, this process was likely responsible for at least three of the four most recent influenza pandemics (Kawaoka et al., 1989; Scholtissek et al., 1978; G. J. D. Smith et al., 2009). Compared to seasonal influenza, the nature of a future influenza pandemic is extremely uncertain: influenza pandemics often circulate freely outside the typical influenza season (Shaman et al., 2011), and severity of past influenza pandemics has varied extremely (Kilbourne, 2006).

Options for influenza control

Vaccines. Influenza vaccines are the most effective tool we currently have in reducing influenza impact (Osterholm, Kelley, Manske, et al., 2012; WHO, 2011a). In addition to preventing

infection, evidence suggests that those who become infected despite vaccination experience less severe disease than those who are unvaccinated (Arriola et al., 2017; Thompson et al., 2018).

That said, influenza vaccines exhibit several significant drawbacks. Because the immune system reacts most strongly to the rapidly evolving viral surface proteins HA and NA (Carrat & Flahault, 2007; Osterholm, Kelley, Manske, et al., 2012), current vaccines are strain-specific, containing three or four influenza viruses each season (Chen et al., 2020; Grohskopf et al., 2019). Coupled with the fact that vaccine production takes six to eight months, this means that vaccine efficacy hinges somewhat on experts' ability to accurately predict the next season's circulating strains several months in advance (Chen et al., 2020; Hay & McCauley, 2018). Mismatch between circulating and vaccine strains contributes to low vaccine efficacy. A 2012 meta-analysis reported an average vaccine efficacy of 59% (Osterholm, Kelley, Sommer, et al., 2012), and a recent analysis calculated overall vaccine efficacy in the United Kingdom (UK) during the 2017-18 season at only 15% (Pebody et al., 2019). Long production times and strain specificity also mean that vaccines are unlikely to be available during the early stages of an emerging pandemic (Wood, 2001). Indeed, the vaccine against the 2009 pandemic strain was not completed until after the outbreak had peaked (Lipsitch et al., 2011).

In addition to vaccine efficacy, duration of protection is also in question; vaccine-induced immunity may begin to wane within one hundred days of vaccination, particularly in the elderly (Castilla et al., 2013). This makes ideal vaccination strategies particularly unclear in the tropics and subtropics, where year-round circulation makes it difficult to know when to vaccinate, and whether to recommend the northern or southern hemisphere vaccine (Caini et al., 2016; Hirve et al., 2016).

Finally, vaccine shortages are relatively common, occurring in some form in six of ten Northern hemisphere seasons between 1999 and 2009 (Osterholm, Kelley, Manske, et al., 2012).

Antiviral drugs. Specific antiviral treatments for influenza also exist, the most common of which are the neuraminidase inhibitors (Bassetti et al., 2019; Lampejo, 2020). Patients treated within 48 hours of symptom onset experience both reduced symptom intensity and disease duration (Fiore et al., 2011), although evidence suggests that severely ill patients may benefit from treatment even after 48 hours (McGeer, 2009). In addition to mitigating disease, treatment with antivirals may help prevent onward transmission of influenza (Ng et al., 2010; Pebody et al., 2011), although the reduction in transmissibility is likely relatively small (Kramer & Bansal, 2015). Neuraminidase inhibitors can also be used prophylactically (Fiore et al., 2011; Lampejo, 2020). Although sporadic cases of resistance occur, typically during treatment, widespread circulation of resistant strains is uncommon (Bassetti et al., 2019; Lampejo, 2020).

Nonpharmaceutical interventions. In addition to vaccines and antiviral drugs, various nonpharmaceutical interventions against influenza are also available. The simplest example is hand washing, an intervention recommended by the WHO regardless of outbreak severity (WHO, 2019). Additionally, it is recommended that those exhibiting symptoms of influenza self-isolate to avoid spreading the virus to others, and wear face masks when in contact with uninfected individuals (WHO, 2019).

In the case of a more severe outbreak, community-level interventions may be enacted, as in the current COVID-19 pandemic. For example, evidence suggests that school closures can be effective in both delaying outbreak peaks and reducing peak intensity, especially when

introduced early in an outbreak (Bin Nafisah et al., 2018). Other social distancing measures, such as workplace closures or limiting mass gatherings, may also be effective, although they are less well studied (Ahmed et al., 2018; Aledort et al., 2007; WHO, 2019). Border screenings and travel restrictions may be put into place, although evidence suggests that these measures lead to only minimal delays in epidemic arrival and peak timing (Cooper et al., 2006; Epstein et al., 2007), and are ineffective once local transmission is established (Lipsitch et al., 2011). Unlike social distancing, these interventions are not recommended by the WHO (WHO, 2019, 2020). However, during more severe outbreaks, countries may turn to these measures regardless, as evidenced by extensive between- and within-country travel restrictions in response to COVID-19 ("Travel Updates," 2020).

Like antivirals, nonpharmaceutical interventions are particularly critical for controlling emerging influenza pandemics in the absence of an effective vaccine (Aledort et al., 2007). However, as observed with both school closures and air travel restrictions, these interventions rapidly lose effectiveness if not implemented in a timely fashion.

Influenza surveillance

At least 60% of countries operate some form of influenza surveillance system (Hay & McCauley, 2018; WHO, 2011a), which may be operational year-round or only during the influenza season (WHO, 2011b, 2014). Many of these are sentinel systems, in which surveillance is conducted systematically at a number of specifically chosen sites (as opposed to universal surveillance) (WHO, 2014). These sites may be outpatient clinics or hospitals, and may include facilities serving the general population or facilities, like pediatric clinics, that serve specific groups (WHO, 2011a). The number and distribution of sentinel sites will depend on the

sociodemographic and climatic characteristics of a country, as well as country capacity (WHO, 2011a, 2014). Broadly, influenza surveillance falls into two categories: epidemiologic and virologic surveillance, each with their own advantages, although the same sentinel systems can be used to collect both types of data.

Epidemiologic surveillance: Epidemiologic (or syndromic) surveillance involves recording data on patients presenting with certain collections of symptoms. The most common syndromes under surveillance are influenza-like illness (ILI) and severe acute respiratory infection (SARI). Although definitions vary, the WHO case definition for ILI requires fever of at least 38°C (100.4°F) and cough with onset within the past 10 days; patients with SARI have all of these symptoms, and require hospitalization (WHO, 2014). Notably, these patients may or may not actually have influenza, as a variety of other infections yield similar symptoms (WHO, 2014). Ideally, sociodemographic data, as well as data on any conditions known to be risk factors for severe influenza, should also be recorded (WHO, 2014). This information can help experts identify outbreak onset, understand the relative severity of an outbreak, and identify groups at risk for severe illness (ECDC, n.d.-a; WHO, 2011b).

Virologic surveillance: Virologic surveillance for influenza consists of the collection and testing of samples from potentially infected patients. This may involve testing of patients identified under epidemiologic surveillance as ILI or SARI cases, ideally in some systematic fashion so as to minimize bias (WHO, 2014). Additional samples may be taken for diagnostic purposes based on the judgement of individual physicians, including those not at sentinel facilities (WHO, 2011b). In addition to testing for influenza, some samples may be further characterized through

subtyping, or may be assessed for antiviral resistance, although these tests are not performed for every sample taken (WHO, 2011a). Similar to epidemiologic surveillance, virologic surveillance can inform estimates of the relative severity of a given outbreak (WHO, 2011a). Furthermore, virologic information helps public health officials understand (sub)type circulation patterns, monitor the emergence of antiviral resistance, and decide upon candidate strains for the next influenza vaccine (ECDC, n.d.-b; WHO, 2011a). Finally, these surveillance systems act as early warning systems in the event of human infection with novel influenza subtypes (Hay & McCauley, 2018; WHO, 2011a).

Ideally, surveillance systems collect both virologic and epidemiologic data, although virologic surveillance is much more established in most regions (WHO, 2014). The WHO's Global Influenza Surveillance and Response System (GISRS), designed to promote global virologic surveillance and virus sharing, has been in place since 1952 (WHO, 2014), and currently consists of over 140 laboratories in 115 countries (WHO, n.d.-a, 2011a), although countries in the tropics remain underrepresented (Ortiz et al., 2009). In contrast, the importance of epidemiologic surveillance has only more recently become widely recognized, in part driven by the inability of many surveillance systems to characterize the severity of the 2009 pandemic (Ortiz et al., 2009; WHO, 2010, 2014). While the WHO has extensive guidelines for establishing sentinel surveillance systems capable of collecting robust and relatively unbiased epidemiologic and virologic data, the ability of countries to adopt these measures varies greatly (WHO, 2014), and most countries do not achieve the standards in full.

Despite these limitations, routine surveillance activities are critical in informing the public health response to influenza outbreaks and pandemics. For example, identification of

season onset allows practitioners to issue at least a general warning to medical professionals about the potential for an upcoming surge in patients, and can help public health officials optimize the timing of vaccination (WHO, 2014), especially in the tropics and subtropics (Hirve et al., 2016). Information on those most severely affected during a given outbreak can also help policy makers prioritize the allocation of vaccines and antivirals, especially in the case of shortages (ECDC, n.d.-a; WHO, 2011b). If particularly high attack rates are noticed, strategies such as permitting over-the-counter dispensing of antivirals could be considered (Hayden, 2004). In sum, high-quality surveillance data facilitate public health response to the unique pressures of a given influenza season or pandemic as the outbreak unfolds.

FluNet and FluID

In addition to data collection, the WHO recognizes the need for effective sharing of influenza surveillance data (WHO, 2014). Although reporting of influenza to the WHO is only required for novel viruses (Hay & McCauley, 2018; WHO, 2016), countries are "strongly encouraged" to report both virologic and epidemiologic surveillance data to the WHO (WHO, 2014). Data reported to the WHO, or else collected from WHO regional databases, are organized into two publicly available databases: FluNet (WHO, n.d.-d) for virologic data, and FluID (WHO, n.d.-c) for epidemiologic data. In accordance with influenza surveillance priorities historically, FluNet has existed since 1997 (WHO, n.d.-d), while FluID was only created in 2009 (Biggerstaff et al., 2019). Ideally, countries submit data weekly; data are then published with a one-week lag.

Potential of influenza forecasts

While the myriad ways in which influenza surveillance data are currently used undoubtedly contribute to a reduction in influenza morbidity and mortality, it should be noted that the majority of the goals laid out by public health organizations like the WHO (WHO, 2014) and the European Centre for Disease Prevention and Control (ECDC) (ECDC, n.d.-a) are reactive in nature: patterns in circulation and severity are observed, and control measures are applied accordingly (Lutz et al., 2019). Of the control measures available against influenza, only vaccination is inherently proactive, but frequent shortages can make it difficult to target vaccine allocation effectively (Bansal et al., 2006; Shimabukuro et al., 2007). Furthermore, even seasonal influenza routinely overwhelms hospitals and emergency departments (Burmeister et al., 2017; Elliot et al., 2008; "Virulent Influenza," 2013), indicating the need for innovative tools and strategies to combat influenza. While there is no doubt that improvements can be made in how public health practitioners react to influenza outbreaks as they unfold, greater reductions in morbidity and mortality may be possible if public health and medical professionals had more time to prepare, rather than reacting in real-time (CDC, n.d.-c).

Forecasting involves using data from the past and present to make an informed prediction of some future outcome of interest (Lutz et al., 2019). Notably, this involves not only a point prediction, but also some measure of forecast certainty. This information is critical, because the level of certainty in an outcome can play just as large a role in driving decision making as the outcome itself. Consider, for example, the forecasts shown in Figure 3. Both forecasts are generated in early December (week 49), and the mean trajectories predict that, during the week of January first, one thousand new cases of influenza are expected. The use of these forecasts, however, or indeed whether a forecast is strongly considered in public health planning, will

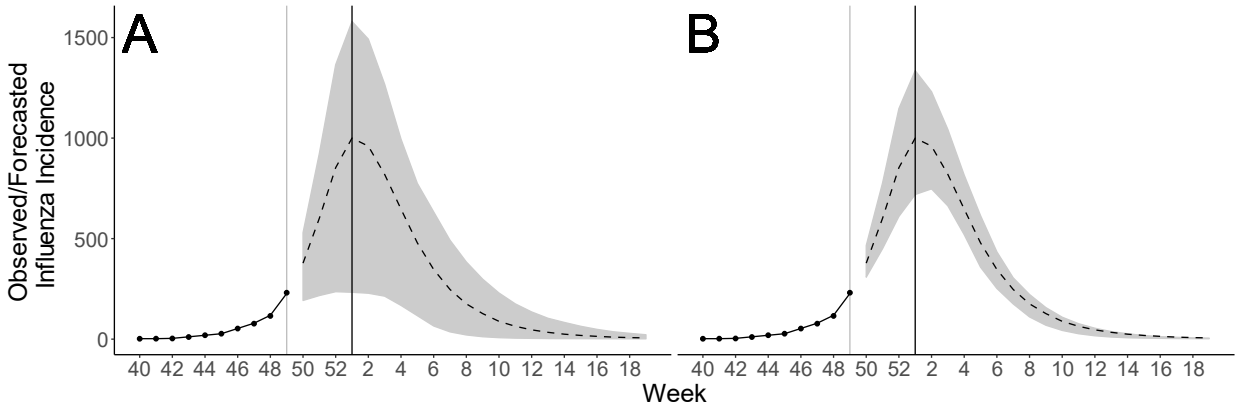


Figure 3. Example influenza forecasts showing identical mean predictions but varying certainty. Points represent weekly data available at Week 49 (gray vertical line), and the solid line shows the fit of the forecasting model to these data. The dotted line represents the mean expected forecast trajectory; the shaded gray areas represent 90% prediction intervals. The week of January first is marked with a black vertical line.

depend on the prediction interval around this prediction. In Figure 3A, this interval is relatively wide, including anywhere from about 200 to 1500 new cases at the peak. In Figure 3B, however, the range is much more narrow, ranging from about 700 to 1300 new cases. While the same mean outcome may be forecast, these two prediction intervals indicate drastically different certainties.

Given sufficient certainty, forecasts could allow public health and medical practitioners to better prepare for an unfolding outbreak or pandemic (CDC, n.d.-c). For example, hospitals in areas where intense influenza activity is predicted could prepare more beds and personal protective equipment, and plan to have more staff on duty. Public health practitioners could use forecasts of upcoming outbreak peaks to encourage timely vaccination, or even to develop more targeted vaccination policies in the case of a severe season or a vaccine shortage. Finally, the general public could use information about upcoming influenza activity to modify their behavior accordingly, perhaps by avoiding large events and crowded areas, thereby reducing influenza transmission.

In this way, skillful forecasting of influenza activity could drive a paradigm shift in how influenza outbreaks are controlled. Rather than waiting for and reacting to increases in influenza activity, communities could prepare for and prevent spikes in morbidity and mortality. Notably, there are already many examples of skillful influenza forecasts (Moss et al., 2017; Ong et al., 2010; Shaman et al., 2013; Shaman & Karspeck, 2012; W. Yang et al., 2015). However, use of these forecasts in practice remains uncommon, in part because forecasts lack the accuracy and certainty needed to support decision making (Biggerstaff et al., 2018; Lipsitch et al., 2011), and in part because effective collaboration between the modelers responsible for forecasting and the public health professionals responsible for policy decisions remains rare (Driedger et al., 2014; Moss et al., 2018).

Public health interest in forecasting

In recent years there has been a substantial push for the further development of influenza forecasting tools among public health organizations. The United States Centers for Disease Control and Prevention (CDC), for example, has held a yearly forecasting challenge since the 2013-14 influenza season (Biggerstaff et al., 2016), which has encouraged the development and improvement of forecasting systems. The forecasting challenge has also acted as an opportunity to practice effective communication of the results of forecasting, both between the modelers producing forecasts and the CDC (CDC, n.d.-a, n.d.-c), and between the CDC and the American public (CDC, n.d.-c). In 2017, the WHO listed the use of models, including forecasts, in public health decision making as a topic of high importance for influenza research in the next five years (WHO, 2017). The WHO's commitment to such research can be observed in the recent development of the Influenza Incidence Analytics Group (IIAG), a global consortium of

researchers working to improve the real-time use of the WHO's FluNet and FluID data (Biggerstaff et al., 2019). The push for the development of more accurate and certain forecasts, in combination with emerging interest from public health organizations, suggests a promising future for the use of influenza forecasts in practical applications.

Aims

This dissertation aims to (1) develop and improve country-level influenza forecasts using publicly-available influenza surveillance data from the WHO; (2) explore and discuss barriers to the use of forecasts in public health practice.

In Chapter 2, we use FluNet and FluID data to generate retrospective forecasts of influenza activity in 64 countries, including 18 in the tropics and subtropics. We assess both the accuracy and the calibration (i.e., the extent to which forecast variability was indicative of outcome accuracy) of these forecasts, and compare forecasts for countries in temperate versus tropical regions. We also compare forecasts by the type of epidemiologic data collected (ILI, SARI, acute respiratory infection (ARI), or pneumonia). Finally, we assess the influence of absolute humidity forcing on forecast accuracy in temperate regions.

In Chapter 3, we develop a metapopulation model of influenza transmission for twelve countries in Europe, by incorporating both commuting and air travel into the model structure. We use this model to generate retrospective forecasts of influenza, and compare our results to forecasts generated for individual countries, as in Chapter 2. We also discuss the need for the collection and reporting of comparable influenza data between countries.

In Chapter 4, we use the forecasting system from Chapter 2 to develop real-time forecasts of influenza activity in 37 countries during the 2017-18 and 2018-19 seasons. We compare the

accuracy of these forecasts, generated using real-time FluNet and FluID data, to that of forecasts generated using data downloaded at the end of the influenza season.

Finally, in Chapter 5, we explore the extent to which public health practitioners in the United States are aware of and use influenza forecasts in their work. We discuss the implications of participants' responses for the future of influenza forecasting; in particular, we emphasize the need for more effective communication between modelers and public health practitioners.

References

Ahmed, F., Zviedrite, N., & Uzicanin, A. (2018). Effectiveness of workplace social distancing measures in reducing influenza transmission: A systematic review. *BMC Public Health*, *18*(1), 518. https://doi.org/10.1186/s12889-018-5446-1

Aledort, J. E., Lurie, N., Wasserman, J., & Bozzette, S. A. (2007). Non-pharmaceutical public health interventions for pandemic influenza: An evaluation of the evidence base. *BMC Public Health*, *7*(1), 208. https://doi.org/10.1186/1471-2458-7-208

Arriola, C., Garg, S., Anderson, E. J., Ryan, P. A., George, A., Zansky, S. M., Bennett, N., Reingold, A., Bargsten, M., Miller, L., Yousey-Hindes, K., Tatham, L., Bohm, S. R., Lynfield, R., Thomas, A., Lindegren, M. L., Schaffner, W., Fry, A. M., & Chaves, S. S. (2017). Influenza Vaccination Modifies Disease Severity Among Community-dwelling Adults Hospitalized With Influenza. *Clinical Infectious Diseases*, *65*(8), 1289–1297. https://doi.org/10.1093/cid/cix468

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, *106*(51), 21484–21489.

Bansal, S., Pourbohloul, B., & Meyers, L. A. (2006). A Comparative Analysis of Influenza Vaccination Programs. *PLoS Medicine*, *3*(10), e387. https://doi.org/10.1371/journal.pmed.0030387

Barreca, A. I., & Shimshack, J. P. (2012). Absolute Humidity, Temperature, and Influenza Mortality: 30 Years of County-Level Evidence from the United States. *American Journal of Epidemiology*, *176*(suppl_7), S114–S122. https://doi.org/10.1093/aje/kws259

Bassetti, M., Castaldo, N., & Carnelutti, A. (2019). Neuraminidase inhibitors as a strategy for influenza treatment: Pros, cons and future perspectives. *Expert Opinion on Pharmacotherapy*, *20*(14), 1711–1718. https://doi.org/10.1080/14656566.2019.1626824

Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I. C.-H., Hickmann, K. S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., Velardi, P., Vespignani, A., & Finelli, L. (2016). Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infectious Diseases*, *16*. https://doi.org/10.1186/s12879-016-1669-x

Biggerstaff, M., Dahlgren, F. S., Fitzner, J., George, D., Hammond, A., Hall, I., Haw, D., Imai, N., Johansson, M. A., Kramer, S., McCaw, J. M., Moss, R., Pebody, R., Read, J. M., Reed, C., Reich, N. G., Riley, S., Vandemaele, K., Viboud, C., & Wu, J. T. (2019). Coordinating the real-time use of global influenza activity data for better public health planning. *Influenza and Other Respiratory Viruses*, *14*(2), 105–110. https://doi.org/10.1111/irv.12705

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. (2018). Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*. https://doi.org/10.1016/j.epidem.2018.02.003

Bin Nafisah, S., Alamery, A. H., Al Nafesa, A., Aleid, B., & Brazanji, N. A. (2018). School closure during novel influenza: A systematic review. *Journal of Infection and Public Health*, *11*(5), 657–661. https://doi.org/10.1016/j.jiph.2018.01.003

Bloom-Feshbach, K., Alonso, W. J., Charu, V., Tamerius, J., Simonsen, L., Miller, M. A., & Viboud, C. (2013). Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PLOS ONE*, *8*(2), e54445. https://doi.org/10.1371/journal.pone.0054445

Bozick, B. A., & Real, L. A. (2015). The Role of Human Transportation Networks in Mediating the Genetic Structure of Seasonal Influenza in the United States. *PLOS Pathogens*, *11*(6), e1004898. https://doi.org/10.1371/journal.ppat.1004898

Bresee, J., & Hayden, F. G. (2013). Epidemic Influenza—Responding to the Expected but Unpredictable. *New England Journal of Medicine*, *368*(7), 589–592. https://doi.org/10.1056/NEJMp1300375

Brockmann, D., & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science*, *342*(6164), 1337–1342.

Brownstein, J. S., Wolfe, C. J., & Mandl, K. D. (2006). Empirical Evidence for the Effect of Airline Travel on Inter-Regional Influenza Spread in the United States. *PLoS Medicine*, *3*(10), e401. https://doi.org/10.1371/journal.pmed.0030401

Burmeister, D., Daya, M., & Sampson, C. (2017). EDs Report Strong Surges in Flu-related Volume, Straining Capacity in Some Regions. *ED Management: The Monthly Update on Emergency Department Management*, *29*(4), 37–40.

Caini, S., Andrade, W., Badur, S., Balmaseda, A., Barakat, A., Bella, A., Bimohuen, A., Brammer, L., Bresee, J., Bruno, A., Castillo, L., Ciblak, M. A., Clara, A. W., Cohen, C., Cutter, J., Daouda, C., de Lozano, C., De Mora, D., Dorji, K., … Global Influenza B Study. (2016). Temporal Patterns of Influenza A and B in Tropical and Temperate Countries: What Are the Lessons for Influenza Vaccination? *PloS One*, *11*(3), e0152310. https://doi.org/10.1371/journal.pone.0152310

Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. *Vaccine*, *25*(39–40), 6852–6862. https://doi.org/10.1016/j.vaccine.2007.07.027

Castilla, J., Martínez-Baz, I., Martínez-Artola, V., Reina, G., Pozo, F., García Cenoz, M., Guevara, M., Moran, J., Irisarri, F., Arriazu, M., Albéniz, E., Ezpeleta, C., Barricarte, A., Primary Health Care Sentinel Networ, C., & Network for Influenza Surveillance, C. (2013). Decline in influenza vaccine effectiveness with time after vaccination, Navarre, Spain, season 2011/12. *Eurosurveillance*, *18*(5). https://doi.org/10.2807/ese.18.05.20388-en

CDC. (n.d.-a). *How CDC Uses Flu Forecasting*. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/weekly/flusight/how-flu-forecasting.htm

CDC. (n.d.-b). *Influenza Type A Viruses*. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/avianflu/influenza-a-virus-subtypes.htm

CDC. (n.d.-c). *Why CDC Supports Flu Forecasting*. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/weekly/flusight/why-flu-forecasting.htm

Charaudeau, S., Pakdaman, K., & Boëlle, P.-Y. (2014). Commuter Mobility and the Spread of Infectious Diseases: Application to Influenza in France. *PLoS ONE*, *9*(1), e83002. https://doi.org/10.1371/journal.pone.0083002

Charu, V., Zeger, S., Gog, J., Bjørnstad, O. N., Kissler, S., Simonsen, L., Grenfell, B. T., & Viboud, C. (2017). Human mobility and the spatial transmission of influenza in the United States. *PLOS Computational Biology*, *13*(2), e1005382. https://doi.org/10.1371/journal.pcbi.1005382

Chen, J.-R., Liu, Y.-M., Tseng, Y.-C., & Ma, C. (2020). Better influenza vaccines: An industry perspective. *Journal of Biomedical Science*, *27*(1), 33. https://doi.org/10.1186/s12929-020-0626-6

Colizza, V., Barthélemy, M., Barrat, A., & Vespignani, A. (2007). Epidemic modeling in complex realities. *Comptes Rendus Biologies*, *330*(4), 364–374. https://doi.org/10.1016/j.crvi.2007.02.014

Cooper, B. S., Pitman, R. J., Edmunds, W. J., & Gay, N. J. (2006). Delaying the international spread of pandemic influenza. *PLoS Medicine*, *3*(6), e212.

Coronavirus travel updates: Which countries have restrictions and FCO warnings in place? (2020, April 13). *The Guardian*. https://www.theguardian.com/travel/2020/mar/24/coronavirus-travel-updates-which-countries-have-restrictions-and-fco-warnings-in-place

Cox, N. J., & Subbarao, K. (2000). Global Epidemiology of Influenza: Past and Present. *Annual Review of Medicine*, *51*(1), 407–421. https://doi.org/10.1146/annurev.med.51.1.407

Crepey, P., & Barthelemy, M. (2007). Detecting Robust Patterns in the Spread of Epidemics: A Case Study of Influenza in the United States and France. *American Journal of Epidemiology*, *166*(11), 1244–1251. https://doi.org/10.1093/aje/kwm266

Dawood, F. S., Iuliano, A. D., Reed, C., Meltzer, M. I., Shay, D. K., Cheng, P.-Y., Bandaranayake, D., Breiman, R. F., Brooks, W. A., Buchy, P., Feikin, D. R., Fowler, K. B., Gordon, A., Hien, N. T., Horby, P., Huang, Q. S., Katz, M. A., Krishnan, A., Lal, R., … Widdowson, M.-A. (2012). Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study. *The Lancet Infectious Diseases*, *12*(9), 687–695. https://doi.org/10.1016/S1473-3099(12)70121-4

De Salazar, P. M., Niehus, R., Taylor, A., Buckee, C. O., & Lipsitch, M. (2020). Identifying Locations with Possible Undetected Imported Severe Acute Respiratory Syndrome Coronavirus 2 Cases by Using Importation Predictions. *Emerging Infectious Diseases*, *26*(7). https://doi.org/10.3201/eid2607.200250

Driedger, S. M., Cooper, E. J., & Moghadas, S. M. (2014). Developing model-based public health policy through knowledge translation: The need for a 'Communities of Practice.' *Public Health*, *128*(6), 561–567. https://doi.org/10.1016/j.puhe.2013.10.009

ECDC. (n.d.-a). *Facts about influenza surveillance*. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/seasonal-influenza/surveillance-and-disease-data/facts

ECDC. (n.d.-b). *Global influenza surveillance and virus sharing*. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/seasonal-influenza/surveillance-and-disease-data/facts-global-surveillance

Elliot, A. J., Cross, K. W., & Fleming, D. M. (2008). Acute respiratory infections and winter pressures on hospital admissions in England and Wales 1990-2005. *Journal of Public Health*, *30*(1), 91–98. https://doi.org/10.1093/pubmed/fdn003

Epstein, J. M., Goedecke, D. M., Yu, F., Morris, R. J., Wagener, D. K., & Bobashev, G. V. (2007). Controlling Pandemic Flu: The Value of International Air Travel Restrictions. *PLoS ONE*, *2*(5), e401. https://doi.org/10.1371/journal.pone.0000401

Fiore, A. E., Fry, A., Shay, D., Gubareva, L., Bresee, J. S., Uyeki, T. M., & Centers for Disease Control and Prevention (CDC). (2011). Antiviral agents for the treatment and chemoprophylaxis of influenza—Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR. Recommendations and Reports: Morbidity and Mortality Weekly Report. Recommendations and Reports*, *60*(1), 1–24.

Galanti, M., Birger, R., Ud-Dean, M., Filip, I., Morita, H., Comito, D., Anthony, S., Freyer, G. A., Ibrahim, S., Lane, B., Matienzo, N., Ligon, C., Rabadan, R., Shittu, A., Tagne, E., & Shaman, J. (2019). Rates of asymptomatic respiratory virus infection across age groups. *Epidemiology and Infection*, *147*, e176. https://doi.org/10.1017/S0950268819000505

Geoghegan, J. L., Saavedra, A. F., Duchêne, S., Sullivan, S., Barr, I., & Holmes, E. C. (2018). Continental synchronicity of human influenza virus epidemics despite climactic variation. *PLOS Pathogens*, *14*(1), e1006780. https://doi.org/10.1371/journal.ppat.1006780

Grohskopf, L. A., Alyanak, E., Broder, K. R., Walter, E. B., Fry, A. M., & Jernigan, D. B. (2019). Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices — United States, 2019–20 Influenza Season. *MMWR. Recommendations and Reports*, *68*(3), 1–21. https://doi.org/10.15585/mmwr.rr6803a1

Hay, A. J., & McCauley, J. W. (2018). The WHO global influenza surveillance and response system (GISRS)-A future perspective. *Influenza and Other Respiratory Viruses*, *12*(5), 551–557. https://doi.org/10.1111/irv.12565

Hayden, F. G. (2004). Pandemic Influenza: Is an Antiviral Response Realistic? *The Pediatric Infectious Disease Journal*, *23*(Supplement), S262–S269. https://doi.org/10.1097/01.inf.0000144680.39895.ce

Herfst, S., Schrauwen, E. J. A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V. J., Sorrell, E. M., Bestebroer, T. M., Burke, D. F., Smith, D. J., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., & Fouchier, R. A. M. (2012). Airborne transmission of influenza A/H5N1 virus between ferrets. *Science (New York, N.Y.)*, *336*(6088), 1534–1541. https://doi.org/10.1126/science.1213362

Hirve, S., Newman, L. P., Paget, J., Azziz-Baumgartner, E., Fitzner, J., Bhat, N., Vandemaele, K., & Zhang, W. (2016). Influenza Seasonality in the Tropics and Subtropics – When to Vaccinate? *PLOS ONE*, *11*(4), e0153003. https://doi.org/10.1371/journal.pone.0153003

Imai, M., Watanabe, T., Hatta, M., Das, S. C., Ozawa, M., Shinya, K., Zhong, G., Hanson, A., Katsura, H., Watanabe, S., Li, C., Kawakami, E., Yamada, S., Kiso, M., Suzuki, Y., Maher, E. A., Neumann, G., & Kawaoka, Y. (2012). Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature*, *486*(7403), 420–428. https://doi.org/10.1038/nature10831

Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., Cohen, C., Gran, J. M., Schanzer, D., Cowling, B. J., Wu, P., Kyncl, J., Ang, L. W., Park, M., Redlberger-Fritz, M., Yu, H., Espenhain, L., Krishnan, A., Emukule, G., … Mustaquim, D. (2018). Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *The Lancet*, *391*(10127), 1285–1300. https://doi.org/10.1016/S0140-6736(17)33293-2

Kaiser, J. (2006). A One-Size-Fits-All Flu Vaccine? *Science*, *312*(5772), 380–382. https://doi.org/10.1126/science.312.5772.380

Kawaoka, Y., Krauss, S., & Webster, R. G. (1989). Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *Journal of Virology*, *63*(11), 4603–4608.

Khan, K., Arino, J., Hu, W., Raposo, P., Sears, J., Calderon, F., Heidebrecht, C., Macdonald, M., Liauw, J., Chan, A., & Gardam, M. (2009). Spread of a Novel Influenza A (H1N1) Virus via Global Airline Transportation. *New England Journal of Medicine*, *361*(2), 212–214. https://doi.org/10.1056/NEJMc0904559

Kilbourne, E. D. (2006). Influenza Pandemics of the 20th Century. *Emerging Infectious Diseases*, *12*(1), 9–14. https://doi.org/10.3201/eid1201.051254

Kramer, S. C., & Bansal, S. (2015). Assessing the use of antiviral treatment to control influenza. *Epidemiology and Infection*, *143*(8), 1621–1631. https://doi.org/10.1017/S0950268814002520

Lampejo, T. (2020). Influenza and antiviral resistance: An overview. *Eur J Clin Microbiol Infect Dis*, 8.

Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D., & Suchard, M. A. (2014). Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathogens*, *10*(2), e1003932. https://doi.org/10.1371/journal.ppat.1003932

Lipsitch, M., Finelli, L., Heffernan, R. T., Leung, G. M., & Redd, S. C. (2011). Improving the Evidence Base for Decision Making During a Pandemic: The Example of 2009 Influenza A/H1N. *Biosecurity and Bioterrorism*, *9*(2), 28.

Liu, M., Zhao, X., Hua, S., Du, X., Peng, Y., Li, X., Lan, Y., Wang, D., Wu, A., Shu, Y., & Jiang, T. (2015). Antigenic Patterns and Evolution of the Human Influenza A (H1N1) Virus. *Scientific Reports*, *5*(1), 14171. https://doi.org/10.1038/srep14171

Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., Fernandez, D., Greene, S. K., Kipshidze, N., Liu, L., Mgbere, O., McHugh, L. A., Myers, J. F., Siniscalchi, A., Sullivan, A. D., West, N., Johansson, M. A., & Biggerstaff, M. (2019). Applying infectious disease forecasting to public health: A path forward using influenza forecasting examples. *BMC Public Health*, *19*(1), 1659. https://doi.org/10.1186/s12889-019-7966-8

McGeer, A. J. (2009). Diagnostic Testing or Empirical Therapy for Patients Hospitalized with Suspected Influenza: What to Do? *Clinical Infectious Diseases*, *48*(S1), S14–S19. https://doi.org/10.1086/591852

Moss, R., Fielding, J. E., Franklin, L. J., Stephens, N., McVernon, J., Dawson, P., & McCaw, J. M. (2018). Epidemic forecasts as a tool for public health: Interpretation and (re)calibration. *Australian and New Zealand Journal of Public Health*, *42*(1), 69–76. https://doi.org/10.1111/1753-6405.12750

Moss, R., Zarebski, A., Dawson, P., & McCAW, J. M. (2017). Retrospective forecasting of the 2010–2014 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology & Infection*, *145*(1), 156–169. https://doi.org/10.1017/S0950268816002053

Ng, S., Cowling, B. J., Fang, V. J., Chan, K. H., Ip, D. K. M., Cheng, C. K. Y., Uyeki, T. M., Houck, P. M., Malik Peiris, J. S., & Leung, G. M. (2010). Effects of Oseltamivir Treatment on Duration of Clinical Illness and Viral Shedding and Household Transmission of Influenza Virus. *Clinical Infectious Diseases*, *50*(5), 707–714. https://doi.org/10.1086/650458

Ng, S., & Gordon, A. (2015). Influenza Burden and Transmission in the Tropics. *Current Epidemiology Reports*, *2*(2), 89–100. https://doi.org/10.1007/s40471-015-0038-4

Ong, J. B. S., Chen, M. I.-C., Cook, A. R., Lee, H. C., Lee, V. J., Lin, R. T. P., Tambyah, P. A., & Goh, L. G. (2010). Real-Time Epidemic Monitoring and Forecasting of H1N1-2009 Using Influenza-Like Illness from General Practice and Family Doctor Clinics in Singapore. *PLoS ONE*, *5*(4), e10036. https://doi.org/10.1371/journal.pone.0010036

Ortiz, J. R., Sotomayor, V., Uez, O. C., Oliva, O., Bettels, D., McCarron, M., Bresee, J. S., & Mounts, A. W. (2009). Strategy to Enhance Influenza Surveillance Worldwide. *Emerging Infectious Diseases*, *15*(8), 1271–1278. https://doi.org/10.3201/eid1508.081422

Osterholm, M. T., Kelley, N. S., Manske, J. M., Ballering, K. S., Leighton, T. R., & Moore, K. A. (2012). *The Compelling Need for Game-Changing Influenza Vaccines*. 160.

Osterholm, M. T., Kelley, N. S., Sommer, A., & Belongia, E. A. (2012). Efficacy and effectiveness of influenza vaccines: A systematic review and meta-analysis. *The Lancet Infectious Diseases*, *12*(1), 36–44. https://doi.org/10.1016/S1473-3099(11)70295-X

Payne, S. (2017). Family Orthomyxoviridae. In *Viruses* (pp. 197–208). Elsevier. https://doi.org/10.1016/B978-0-12-803109-4.00023-4

Pebody, R., Djennad, A., Ellis, J., Andrews, N., Marques, D. F. P., Cottrell, S., Reynolds, A. J., Gunson, R., Galiano, M., Hoschler, K., Lackenby, A., Robertson, C., O'Doherty, M., Sinnathamby, M., Panagiotopoulos, N., Yonova, I., Webb, R., Moore, C., Donati, M., … Zambon, M. (2019). End of season influenza vaccine effectiveness in adults and children in the United Kingdom in 2017/18. *Eurosurveillance*, *24*(31). https://doi.org/10.2807/1560-7917.ES.2019.24.31.1800488

Pebody, R., Harris, R., Kafatos, G., Chamberland, M., Campbell, C., Nguyen-Van-Tam, J. S., McLean, E., Andrews, N., White, P. J., Wynne-Evans, E., Green, J., Ellis, J., Wreghitt, T., Bracebridge, S., Ihekweazu, C., Oliver, I., Smith, G., Hawkins, C., Salmon, R., … Watson, J. M. (2011). Use of Antiviral Drugs to Reduce Household Transmission of Pandemic (H1N1) 2009, United Kingdom1. *Emerging Infectious Diseases*, *17*(6), 990–999. https://doi.org/10.3201/eid/1706.101161

Putri, W. C. W. S., Muscatello, D. J., Stockwell, M. S., & Newall, A. T. (2018). Economic burden of seasonal influenza in the United States. *Vaccine*, *36*(27), 3960–3966. https://doi.org/10.1016/j.vaccine.2018.05.057

Scholtissek, C., Rohde, W., Von Hoyningen, V., & Rott, R. (1978). On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology*, *87*(1), 13–20. https://doi.org/10.1016/0042-6822(78)90153-8

Shaman, J., Goldstein, E., & Lipsitch, M. (2011). Absolute Humidity and Pandemic Versus Epidemic Influenza. *American Journal of Epidemiology*, *173*(2), 127–135. https://doi.org/10.1093/aje/kwq347

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, *109*(50), 20425–20430. https://doi.org/10.1073/pnas.1208772109

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3837

Shaman, J., & Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3243–3248. https://doi.org/10.1073/pnas.0806852106

Shimabukuro, T. T., Wortley, P. M., Bardenheier, B., Bresnitz, E. A., DeBlois, A. M., Hahn, C. G., & Mangione, E. J. (2007). Survey of State Practices during the 2004–2005 Influenza Vaccine Shortage. *Public Health Reports*, *122*(3), 311–318. https://doi.org/10.1177/003335490712200304

Smith, D. J. (2004). Mapping the Antigenic and Genetic Evolution of Influenza Virus. *Science*, *305*(5682), 371–376. https://doi.org/10.1126/science.1097211

Smith, G. J. D., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., Ma, S. K., Cheung, C. L., Raghwani, J., Bhatt, S., Peiris, J. S. M., Guan, Y., & Rambaut, A. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*, *459*(7250), 1122–1125. https://doi.org/10.1038/nature08182

Tamerius, J., Shaman, J., Alonso, W. J., Bloom-Feshbach, K., Uejio, C. K., Comrie, A., & Viboud, C. (2013). Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLoS Pathogens*, *9*(3), e1003194. https://doi.org/10.1371/journal.ppat.1003194

Tamura, S., Tanimoto, T., & Kurata, T. (2005). Mechanisms of broad cross-protection provided by influenza virus infection and their application to vaccines. *Japanese Journal of Infectious Diseases*, *58*(4), 195–207.

Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: The mother of all pandemics. *Rev Biomed*, *17*, 69–79.

Thompson, M. G., Pierse, N., Sue Huang, Q., Prasad, N., Duque, J., Claire Newbern, E., Baker, M. G., Turner, N., & McArthur, C. (2018). Influenza vaccine effectiveness in preventing influenza-associated intensive care admissions and attenuating severe disease among adults in New Zealand 2012–2015. *Vaccine*, *36*(39), 5916–5925. https://doi.org/10.1016/j.vaccine.2018.07.028

Viboud, C., Bjørnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., & Grenfell, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, *312*(5772), 447–451.

Virulent influenza fills EDs across the country, prompting hospitals to launch emergency plans. (2013). *ED Management: The Monthly Update on Emergency Department Management*, *25*(3), 25–28.

WHO. (n.d.-a). *Global Influenza Surveillance and Response System (GISRS)*. World Health Organization. https://www.who.int/influenza/gisrs_laboratory/en/

WHO. (n.d.-b). *Influenza (Seasonal)*. WHO. https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)

WHO. (n.d.-c). *WHO | FluID - a global influenza epidemiological data sharing platform*. WHO. https://www.who.int/influenza/surveillance_monitoring/fluid/en/

WHO. (n.d.-d). *WHO | FluNet*. WHO. https://www.who.int/influenza/gisrs_laboratory/flunet/en/

WHO. (2010). *Surveillance Recommendations for Member States in the Post Pandemic Period*. World Health Organization. https://www.who.int/csr/resources/publications/swineflu/surveillance_post_pandemic_20100812/en/

WHO. (2011a). *Manual for the laboratory diagnosis and virological surveillance of influenza*. World Health Organization. https://www.who.int/influenza/gisrs_laboratory/manual_diagnosis_surveillance_influenza/en/

WHO. (2011b). *WHO Regional Office for Europe guidance for sentinel influenza surveillance in humans*. http://www.euro.who.int/__data/assets/pdf_file/0020/90443/E92738.pdf

WHO. (2014). *Global Epidemiological Surveillance Standards for Influenza*. World Health Organization. https://www.who.int/influenza/resources/documents/influenza_surveillance_manual/en/

WHO. (2016). *International Health Regulations (2005): Third Edition*. https://www.who.int/ihr/publications/9789241580496/en/

WHO. (2017). *WHO Public Health Research Agenda for Influenza (2017 Update)*. https://www.who.int/influenza/resources/research/publication_research_agenda_2017/en/

WHO. (2018a). *Fact sheet: Global Influenza Surveillance and Response System (GISRS)*. https://www.who.int/influenza/gisrs_laboratory/updates/GISRS_one_pager_2018_EN.pdf

WHO. (2018b). *Influenza (Avian and other zoonotic)*. World Health Organization. https://www.who.int/en/news-room/fact-sheets/detail/influenza-(avian-and-other-zoonotic)

WHO. (2019). *Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza*. https://www.who.int/influenza/publications/public_health_measures/publication/en/

WHO. (2020). *Updated WHO recommendations for international traffic in relation to COVID-19 outbreak*. World Health Organization. https://www.who.int/news-room/articles-detail/updated-who-recommendations-for-international-traffic-in-relation-to-covid-19-outbreak

Wille, M., & Holmes, E. C. (2019). The Ecology and Evolution of Influenza Viruses. *Cold Spring Harbor Perspectives in Medicine*, a038489. https://doi.org/10.1101/cshperspect.a038489

Wood, J. M. (2001). Developing vaccines against pandemic influenza. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1416), 1953–1960. https://doi.org/10.1098/rstb.2001.0981

Yaari, R., Katriel, G., Huppert, A., Axelsen, J. B., & Stone, L. (2013). Modelling seasonal influenza: The role of weather and punctuated antigenic drift. *Journal of The Royal Society Interface*, *10*(84), 20130298–20130298. https://doi.org/10.1098/rsif.2013.0298

Yang, L., Ma, S., Chen, P. Y., He, J. F., Chan, K. P., Chow, A., Ou, C. Q., Deng, A. P., Hedley, A. J., Wong, C. M., & Peiris, J. S. M. (2011). Influenza associated mortality in the subtropics and tropics: Results from three Asian cities. *Vaccine*, *29*(48), 8909–8914. https://doi.org/10.1016/j.vaccine.2011.09.071

Yang, W., Cowling, B. J., Lau, E. H. Y., & Shaman, J. (2015). Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology*, *11*(7), e1004383. https://doi.org/10.1371/journal.pcbi.1004383

Chapter 2

Development and validation of influenza forecasting for 64 temperate and tropical countries

Sarah C Kramer[1] and Jeffrey Shaman[1]

Affiliations:
1. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, 722 W 168th Street, New York, New York, 10032, United States of America

Acknowledgments:
We thank Wan Yang for helpful discussion and input regarding modeling and forecasting influenza in tropical and subtropical regions.

Published: Kramer, S.C. & Shaman, J. 2019. Development and validation of influenza forecasting for 64 temperate and tropical countries. *PLoS Comput Biol, 15*(2).

Abstract

Accurate forecasts of influenza incidence can be used to inform medical and public health decision-making and response efforts. However, forecasting systems are uncommon in most countries, with a few notable exceptions. Here we use publicly available data from the World Health Organization to generate retrospective forecasts of influenza peak timing and peak intensity for 64 countries, including 18 tropical and subtropical countries. We find that accurate and well-calibrated forecasts can be generated for countries in temperate regions, with peak timing and intensity accuracy exceeding 50% at four and two weeks prior to the predicted epidemic peak, respectively. Forecasts are significantly less accurate in the tropics and subtropics for both peak timing and intensity. This work indicates that, in temperate regions around the world, forecasts can be generated with sufficient lead time to prepare for upcoming outbreak peak incidence.

Author Summary

Influenza is responsible for an estimated 3-5 million cases and 300-650,000 deaths each year worldwide. If produced early enough, accurate forecasts of influenza activity could guide public health practitioners and medical professionals in preparing for an outbreak, reducing the subsequent morbidity and mortality. For example, hospitals could use these forecasts to determine how many beds will be needed when an outbreak is most intense. Despite this potential impact, influenza forecasts are primarily generated for the United States, with forecasts for other countries being comparatively rare. Here, we use publicly available influenza data to forecast influenza activity in 64 countries. We find that accurate forecasts can be produced several weeks before the outbreak's peak in temperate countries, where influenza outbreaks

occur regularly during the winter. Forecast accuracy is lower in the tropics and subtropics, where outbreaks occur more sporadically. Overall, our results suggest that forecasts have potential as an important public health tool in many countries, not only in the US.

Introduction

Forecasting is an important tool in a number of fields, including weather and climate (Gneiting & Raftery, 2005; Zebiak et al., 2015; Zebiak & Cane, 1987), agriculture (FAO, 2016; Newlands et al., 2014), air quality (Debry & Mallet, 2014; Gaubert et al., 2014), and consumer activity (Chen & Lu, 2017; Choi et al., 2014; Mccarthy et al., 2006). When operationalized for use in real time, predictions from probabilistic forecasts can be used in decision-making to inform, for example, emergency food aid allocation (Newlands et al., 2014) or profit maximization (Chen & Lu, 2017). Recently, forecasting systems have also been developed for a range of infectious diseases of high public health concern, including influenza (Biggerstaff et al., 2016; Hickmann et al., 2015; Moss et al., 2016, 2017; Nsoesie et al., 2014; Ong et al., 2010; Pei et al., 2018; Shaman et al., 2013; Shaman & Karspeck, 2012; Yang et al., 2015), norovirus (Held et al., 2017), dengue (Adde et al., 2016; Johnson et al., 2017; Reich et al., 2016; Shi et al., 2015), Ebola (Camacho et al., 2015; Funk et al., 2018; Meltzer et al., 2014; Shaman et al., 2014), and, most recently, Zika (Chowell et al., 2016; Huff et al., 2016).

The ability to generate accurate, real-time forecasts of infectious disease activity has important implications for public health. Currently, response to infectious disease outbreaks is primarily reactive: medical and public health professionals attempt to deal with unexpected spikes of disease incidence as they occur. By providing information on when an outbreak is expected to peak and how many cases are expected at that peak, forecasts have the potential to create a paradigm shift in infectious disease control and public health decision-making. For example, hospitals expecting a patient surge might ensure that adequate resources are available, avoiding bed and staff shortages.

Seasonal influenza produces annual wintertime outbreaks in temperate regions, as well as sporadic outbreaks throughout the year in the tropics and subtropics (Bloom-Feshbach et al., 2013; Ng & Gordon, 2015). The World Health Organization (WHO) estimates that influenza causes about 300,000-650,000 deaths and 3-5 million cases of severe illness each year (WHO, n.d.-a). To date, forecasts of influenza activity in the United States have been generated and operationalized (Biggerstaff et al., 2016; Shaman et al., 2013). However, while influenza forecasts have been generated for countries outside the US (Moss et al., 2016, 2017; Nishiura, 2011; Ong et al., 2010; Viboud et al., 2003; Yang et al., 2015), these efforts are less numerous, and many countries have been ignored entirely. The tropics and subtropics are particularly neglected, with forecasts attempted for only Hong Kong (Yang et al., 2015) and Singapore (Ong et al., 2010). This is true despite evidence suggesting that influenza burden in the tropics is similar to that in temperate regions (Ng & Gordon, 2015).

The WHO collects influenza data year-round from several member states around the world. To our knowledge, no influenza forecasts have yet been generated using these data. Given differences in data collection procedures by country, and the importance of high data quality for generating accurate forecasts, whether these data can be used to generate accurate forecasts remains an open question. Here, we explore the following research questions: 1) Can the WHO data be used to generate accurate and well-calibrated retrospective forecasts at the country level?; 2) Does forecast accuracy significantly differ between temperate and tropical regions?; and 3) What factors are associated with substantial changes in forecast accuracy within both temperate and tropical regions? Based on past work, we expect that forecasting will be feasible in all regions, but that forecast accuracy will be substantially higher in temperate regions.

Materials and Methods

Influenza Data

Influenza syndromic and virologic data were obtained from WHO's FluID (WHO, n.d.-b) and FluNet (WHO, n.d.-c) web-tools, respectively. Briefly, these systems contain aggregated influenza data from WHO member states, which are either submitted by member states directly or downloaded by the WHO from existing regional databases. Good quality (see S1 Text) syndromic and virologic data were available for at least one season from 64 countries, primarily in Europe and North America (see Figure 1; S1 Figure). Countries were classified as temperate or tropical based on both their latitude and whether they demonstrated seasonal or sporadic influenza dynamics (see S1 Text). Overall, eighteen countries were classified as tropical, and three (Australia, New Zealand, and Chile) were located in the southern temperate region.

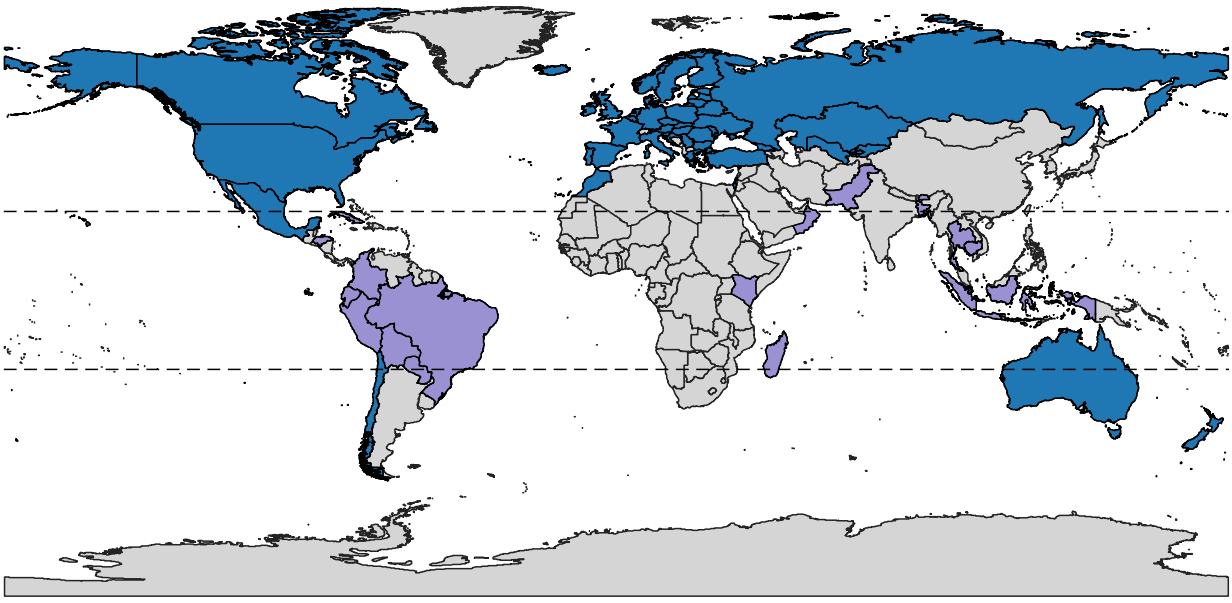


Figure 1. Countries with good quality syndromic and virologic data for at least one season. The black dotted lines demarcate the boundaries of the tropics. Countries classified as temperate are shaded in blue, and countries classified as tropical are shaded in purple. Of 64 countries, 18 were classified as tropical, and 3 as southern temperate.

FluID data include diagnostic counts of influenza-like illness (ILI), acute respiratory infection (ARI), severe acute respiratory infection (SARI), and pneumonia, with different countries preferentially reporting different data types (see S1 Text for additional information). Because these data contain no information on laboratory testing, counts include both patients infected with influenza and patients infected by other pathogens that lead to similar signs and symptoms. To adjust for this lack of specificity, we use FluNet data, which includes the total number of tests performed for influenza and the number positive for influenza. Specifically, we multiply the syndromic case counts from the FluID tool by the proportion of tests positive for influenza in that same country during a given week. This calculation eliminates out-of-season syndromic cases that are unlikely to be due to influenza. Further, as the model used in this study (described below) simulates the transmission of a single pathogen, the removal of incidence due to non-influenza illness increases agreement between model input (data) and output. We refer to the resulting measures as ILI+, ARI+, SARI+, or pneumonia+, or, more broadly, syndromic+.

For this study, we focused specifically on seasonal influenza outbreaks, and excluded the 2009 pandemic from the main analysis. While pandemic outbreaks often produce a strong incidence signal that is forecastable (Ong et al., 2010), they typically appear out-of-season in temperate regions. Seasonal influenza outbreaks, on the other hand, occur with enough frequency that, even in the tropics, where outbreak timing is less regular, future epidemics are almost certain to occur within the year. To maintain a consistent forecasting approach, we therefore focus on seasonal influenza. We present results from forecasting the 2009 pandemic alone, as well as associated methods, in S1 Text and S19 Figure.

In addition, individual seasons were removed from the final dataset if: a) five or more consecutive weeks of data were missing near the outbreak peak (n = 2); b) the season consisted

of fewer than 5 non-NA and non-zero data points (n = 2); c) the total attack rate of the season was less than 5% that of the largest outbreak (in other words, if case counts were unrealistically low; n = 5); d) data collection began or ended at the outbreak peak (n = 4); or e) no consecutive weeks of data were available (in other words, data were only available every other week; n = 1). We also removed data from 2010-11 in Mexico due to the continued disruption of typical seasonal patterns by the 2009 pandemic. Individual data points were removed if they occurred outside of the influenza season (as defined below under "Delineation of Influenza Seasons") and were greater than 50% of the maximum value for the country over all seasons (n = 1). In total, 15 individual seasons were removed from the dataset, and 64 countries remained. In temperate regions, data were available for between one and seven seasons for each country for a total of 289 seasons. A complete list of countries and seasons used for forecasting can be found in S1 and S2 Tables. Note that, for the seasonal forecasts, we began fitting tropical data at week 40 of 2010.

Humidity Data

Data on absolute humidity were obtained from NASA's Global Land Data Assimilation System (GLDAS), which uses both observed data and modeling techniques to produce high-resolution surface meteorological data (Rodell, n.d.). Data were available every three hours at a spatial resolution of 1°x1° for the years 1989-2008. Data from each grid cell were aggregated to the daily level, and anomalous records were identified by visual inspection and removed. Then, climatologies for each grid cell were generated by averaging daily specific humidity on each of 365 days across twelve to twenty years, depending on the amount of anomalous data removed. Finally, climatologies were aggregated to the country level by averaging the climatologies for all

grid cells lying within a country, weighted by the proportion of the grid cell situated within the country in question. A more detailed description of how the humidity data were processed can be found in S1 Text.

Delineation of Influenza Seasons

The influenza season in temperate regions of the northern hemisphere is modeled as beginning in week 40 and ending in week 20 of the following year (*Flu News Europe*, n.d.). We shift these values by one half-year for temperate regions of the southern hemisphere; thus, the influenza season begins in week 14 and continues until week 46.

For tropical regions, where consistent seasonality in influenza infection patterns is not observed, the above methods are not sufficient. Individual outbreaks are instead identified using methodology previously described in Yang et al. (2015). Briefly, outbreak onsets are defined as the first of three consecutive weeks where ILI+ rates exceeded the 33rd percentile of non-zero ILI+ values across all available data for a country. The end of an outbreak is defined as the first of two consecutive weeks below this threshold. To ensure that sporadic spikes in influenza are not counted, we remove any outbreaks where ILI+ counts never exceeded three times its respective onset threshold value.

Retrospective Forecast Generation

Country-level retrospective forecasts are developed using a model-data assimilation system consisting of: (1) influenza observations, as described above, (2) a model of influenza transmission, and (3) a filter to assimilate observations and optimize model simulation and

ensemble forecast. The final two components are described here. These components differed slightly for temperate and tropical regions, and are therefore described separately.

Temperate Regions

SIRS Model: We model influenza transmission in temperate regions using a compartmental, humidity-forced Susceptible-Infected-Recovered-Susceptible (SIRS) model, in which members of the model population move through compartments according to the following equations:

$$\frac{dS}{dt} = \frac{N - S - I}{L} - \frac{\beta(t)IS}{N} - \alpha$$
$$\frac{dI}{dt} = \frac{\beta(t)IS}{N} - \frac{I}{D} + \alpha \quad (1)$$

where N is the total model population size, set arbitrarily to 100,000 for all countries; S and I are the total number of people susceptible and infected, respectively; t is time in days; $\beta(t)$ is the transmission rate at time t; D is the mean infectious period; L is the average duration of immunity before recovery; and α represents the rate of influenza importation from outside the model population, here set to 0.1, or 1 case per 10 days (Shaman & Karspeck, 2012). The basic reproductive number (R_0), a key parameter in infectious disease epidemiology representing the average number of secondary infections arising from a single primary infection in a fully susceptible population, is related to $\beta(t)$ and D by the expression $R_0(t) = \beta(t)D$.

Daily specific humidity modulates $R_0(t)$ as follows:

$$R_0(t) = e^{-180q(t) + \ln(R_{0_{max}} - R_{0_{min}})} + R_{0_{min}} \quad (2)$$

where R_{0max} is the maximum daily basic reproductive number, R_{0min} is the minimum daily basic reproductive number, and *q(t)* is the specific humidity on day *t* (Shaman & Karspeck, 2012). We set *a* equal to -180, based on laboratory regression of influenza virus survival on specific humidity (Shaman & Kohn, 2009). Absolute humidity has been shown to increase influenza survival and transmission in laboratory experiments (Shaman & Kohn, 2009), and model studies indicate that lower absolute humidity during the winter is a significant driver of influenza seasonality in temperate regions (Shaman et al., 2010). Similar models have been used to forecast influenza at the city and state level in the US (Shaman et al., 2010, 2013; Shaman & Karspeck, 2012), and previous work has shown that inclusion of absolute humidity forcing significantly improves forecast performance (Shaman et al., 2017).

Data Assimilation Methods: The above model is fit to the syndromic+ data using the Ensemble Adjustment Kalman Filter (EAKF), a data assimilation method used in weather forecasting (Anderson, 2001). In practice, we randomly initialize an ensemble of simulations (see Forecast Generation below for details) that are then integrated forward per the model equations. At each observation the integration is halted and the ensemble observed state is updated using the EAKF algorithm and that observation, per Bayes Rule:

$$p(X_t|O_{1:t}) \propto p(X_t|O_{1:(t-1)}) \cdot p(O_t|X_t) \quad (3)$$

where $p(X_t|O_{1:(t-1)})$ is the prior distribution of the observed model state (here, the number of newly infected individuals) given all observations prior to time *t*, $p(O_t|X_t)$ is the likelihood of the observation at time *t* given the model state at time *t*, and $p(X_t|O_{1:t})$ is the posterior distribution of the model state given all observations up to and including time *t*. The probability

of the model state is based on the distribution of the ensemble of simulations. Unobserved state variables and parameters (*S*, R_{0max}, R_{0min}, *D*, and *L*) are updated according to cross-ensemble covariability with the observed model state. More details on the EAKF's implementation can be found in S1 Text, as well as in the literature (Anderson, 2001; Pei et al., 2018; Shaman & Karspeck, 2012; Yang et al., 2014).

Forecast Generation: A forecast for week *t* is produced by first iteratively fitting the ensemble of simulations to local observations from the beginning of the season up to and including week *t*, and then integrating the ensemble until the end of the epidemic period using the final inferred states and parameters from the training period (i.e. the posterior at week *t*). This process is repeated throughout the season for weeks 44 through 69 in the northern hemisphere, and weeks 18 through 43 in the southern hemisphere. Thus each ensemble forecast assimilates 5 to 30 weeks of training data. Prior to simulation and forecast, initial values of states and parameters for each ensemble member are randomly selected using Latin hypercube sampling from ranges previously reported ($1.3 \leq R_{0_{max}} \leq 4.0, 0.8 \leq R_{0_{min}} \leq 1.2, 1.5 \leq D \leq 7.0, 365 \leq L \leq 3650$) (Shaman & Karspeck, 2012). In order to account for any stochastic effects during this initialization 5 separate 300-member ensembles were initialized and used to generate forecasts for each location and season. The average results over all ensembles are reported. Variance within an ensemble permits assessment of forecast uncertainty (Shaman & Karspeck, 2012).

Tropical Regions

For the most part, the procedure used to generate retrospective forecasts in the tropics is similar to that used in temperate zones. Differences are described briefly here.

SIRS Model: Because the relationship between absolute humidity and influenza incidence is less clearly understood in the tropics (Deyle et al., 2016; Shaman et al., 2010; Yang et al., 2018), and because humidity data quality in the tropics is poor, we use a simplified model for these countries that does not incorporate absolute humidity forcing. Here, R_0 is defined simply as βD, and neither β nor R_0 change over time. Thus, one less parameter (R_0 vs. R_{0max} and R_{0min}) is fit by our model-data assimilation system when simulating influenza transmission in the tropics. Initial values of R_0 range from 0.8 to 2.2.

Data Assimilation Methods: Because influenza does not exhibit a coherent seasonal pattern in the tropics, model fitting cannot be performed as described above for temperate regions. Rather, fitting is performed continuously, beginning with the first available observation (as early as October 2010) and ending with the last, as described in Yang et al. (2015).

Forecast Generation: Because the duration of influenza outbreaks in the tropics cannot be known in real time, forecasts are not run through the end of an outbreak period, as in temperate countries. Rather, forecasts for a given week are run 40 weeks into the future. As in temperate regions, we perform 5 simulations of 300 ensemble members each.

Choice of Scaling Factors

As described above, model output represents true influenza incidence per 100,000 population. Our data, on the other hand, are obtained by multiplying nonspecific syndromic data by influenza positivity rates among those who actively seek medical care. Furthermore, the

majority of countries included in the WHO data provide no information on the total number of patients seen or the size of the catchment areas from which data were obtained. Thus, our data represent counts, not rates. In order to properly use the EAKF as described above, we must therefore first scale the data such that they are compatible with the model-simulated state space. In effect, the scaling factors map the observed syndromic+ data to the model state space. Scaled data, thus, represent the estimated number of syndromic+ cases per 100,000 population, and can be used for data assimilation. Model output—the simulations and forecasts—can then be scaled back to their original units (e.g. ARI+) for use by individual country public health departments.

Our previous work has shown that SIRS simulations perform optimally when 15-50% of a model population of 100,000 is infected over the course of a modeled epidemic. Therefore, scaling values, γ, for each country were determined by first calculating the range of scaling values yielding a total attack rate between 15% and 50% for each season, i, ([$\gamma_{15,\,i}$, $\gamma_{50,\,i}$]), then choosing a single country-specific scaling value based on the following rule:

$$\gamma = \begin{Bmatrix} if\ \exists\ \gamma\ \in\ \mathbb{R}:\ \gamma_{15,i} < \gamma < \gamma_{50,i}\ \forall\ i: & max_{i=1}^{n}(\gamma_{15,i}) \\ else: & min_{i=1}^{n}(\gamma_{50,i}) \end{Bmatrix} \quad (4)$$

Although forecasts in the tropics were run continuously rather than by season, scaling factors for tropical countries were determined similarly using influenza outbreaks as identified under "Delineation of Influenza Seasons" above.

Scaling values were allowed to vary by country, but not by season: that is, for each country, a single scaling value was chosen and used in retrospective forecasts of all available seasons. As scaling factors are controlling for differences in rates of seeking medical attention, size of the catchment area from which influenza data are collected, and overall population size by country, they vary substantially, from 0.004 in Mexico to 374 in Peru.

Forecast Accuracy and Comparison

Forecasts were evaluated based on their ability to accurately predict outbreak peak timing (the week with the highest number of influenza cases), peak intensity (the number of influenza cases at the peak), and onset timing (the first of three consecutive weeks with influenza activity over some baseline value). Onset baseline values were chosen as 500 simulated cases for temperate countries, and 300 cases for tropical countries (see S1 Text). A forecast was considered accurate for peak timing and onset timing if the predicted value was within one week of the observed, and for peak intensity if the predicted influenza case count was within 25% of the observed. These thresholds, particularly the 1 week cutoff for peak timing accuracy, have been routinely used both in our past work (Kandula et al., 2017; Morita et al., 2018; Moss et al., 2016, 2017; Pei et al., 2018; Shaman et al., 2017; Yang et al., 2014) and in evaluating forecasts submitted to the Centers for Disease Control and Prevention's (CDC) Predict the Influenza Season Challenge (Biggerstaff et al., 2016), allowing for comparison between the results of this work and past work. If the mode predicted onset timing is NA (no outbreak), predicted peak timing, peak intensity, and onset timing were set to NA, and the forecast was removed from consideration.

Forecast accuracy was compared for temperate vs. tropical regions, as well as within temperate regions by hemisphere, region, data type, season, and scaling, and within the tropics by region, data type, and scaling. Because, in real time, the actual time to peak is unknown, we evaluated forecast accuracy by predicted lead time (i.e. the difference between the week at which a forecast is initiated and predicted peak timing). For most analyses, forecast accuracy was assessed at predicted lead weeks -6 to 4 (i.e. six weeks before the predicted peak through four weeks after). Comparisons were made for each individual variable using generalized estimating

equations (see S1 Text for more details). To assess whether the effects of the explanatory factors change over time, GEE models were also run restricting the data to either before or after the predicted peak. Seasons with no identified onset (in other words, where no outbreak occurred) were removed before analyzing forecast accuracy. Additionally, because individual outbreaks within tropical countries are identified during the forecasting process, and therefore were not checked for quality previously, outbreaks where a) five or more consecutive weeks of data were missing; or b) data collection for an outbreak began at the outbreak peak were removed from tropical countries' results before GEEs were run.

To assess the impact of including humidity forcing in the temperate models, we generated an additional set of forecasts for the temperate regions, this time without including humidity forcing in the model structure (see S1 Text). This resulted in two distinct forecasts for each country, season, start week, and run: one incorporating humidity data and one not. In order to fully take advantage of this paired design, forecast accuracy was compared by observed lead week using the exact binomial test. Because individual comparisons were made for each lead week, we applied a Bonferroni correction and considered differences to be statistically significant when p-values were less than 0.0045 ($p = 0.05 / 11$). Unlike in previous analyses, rather than removing forecasts predicting no onset, we considered these forecasts to be "inaccurate." This was done to avoid ignoring pairs of forecasts where one failed to recognize an oncoming outbreak but the other accurately predicted peak timing or intensity.

Sensitivity analyses were performed to test how forecast accuracy changes as a function of EAKF observational error variance, onset baseline value, scaling, and accuracy metric. Findings from these sensitivity analyses broadly agree with the results presented here (results in S1 Text).

Results

Influenza Data

Retrospective forecasts were performed using syndromic+ data from 64 countries, of which 18 were classified as tropical. In the temperate regions, data were available for between 2 and 7 seasons, with each country contributing an average of 6 seasons of data (data in S2 Table). In the tropics, data were available for between 29 and 345 weeks (mean = 166 weeks; median = 140 weeks). In the northern temperate region, onset timing occurred between weeks 45 and 64, and peak timing occurred between weeks 48 and 67. In the southern temperate region, these values were weeks 23 and 33 for onset timing and 29 and 38 for peak timing.

Forecast Feasibility

Overall, we found that accurate forecasts of both peak timing and peak intensity for influenza outbreaks are possible using publicly available WHO data. In temperate regions, we were able to develop country-level, retrospective forecasts that exceeded 50% accuracy for peak timing (i.e., 50% of forecasts predicted peak timing within one week of the observed value) up to four weeks before the predicted peak, and for peak intensity (within 25% of the observed value) two weeks before the predicted peak. Forecasts exceeded 75% accuracy for peak timing one week before the predicted peak, and for peak intensity at the predicted peak week (Figure 2A). Forecast accuracy was lower in the tropics, never exceeding 50% for either peak timing or peak intensity (Figure 2B). As expected (Moss et al., 2016; Ong et al., 2010; Shaman et al., 2013; Shaman & Karspeck, 2012; Yang et al., 2015), forecast accuracy varied as a function of lead time, with forecasts near and after the forecasted peak typically performing better than forecasts generated several weeks before the peak. Similar patterns were seen by observed lead time,

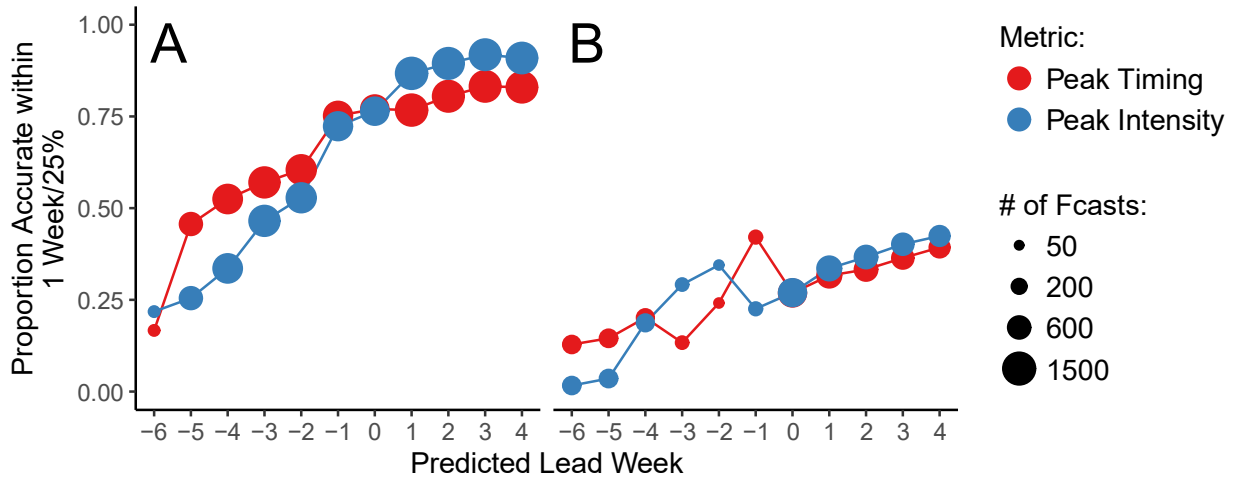


Figure 2. Peak timing and intensity forecast accuracy by predicted lead week.
(A) Forecast accuracy in temperate regions. (B) Forecast accuracy in tropical regions. Peak timing accuracy is shown in red, and peak intensity in blue. The size of the circles represents the number of forecasts generated at a particular lead week.

although tropical forecast accuracy was much higher after the observed peak, exceeding 70% (results in S3 Figure). Broadly, these results remained consistent after altering the cutoff point at which forecasts were considered accurate (S11 Figure), and when correlation coefficients and symmetric mean absolute percentage error (sMAPE) over the entire forecast period were assessed (S12 Figure), although forecast accuracy assessed using sMAPE was comparable between temperate and tropical regions.

Table 1. Onset timing accuracy and number of forecasts predicting any onset by predicted onset week.

Lead Week		-6	-5	-4	-3	-2	-1	0	1	2	3	4
Temperate (w/ humidity)	Accuracy	-	-	-	-	-	41.4%	87.0%	95.6%	95.7%	95.5%	95.2%
	# of Fcasts	0	0	0	0	0	29	1076	1257	1335	1319	1320
Tropical	Accuracy	13.3%	12.7%	11.1%	10.7%	6.7%	50.0%	47.2%	67.3%	70.7%	69.1%	69.1%
	# of Fcasts	165	267	305	290	104	14	339	284	300	285	285

For both temperate and tropical regions, forecasts of outbreak onset timing showed high accuracy post-onset, but forecasts were rarely generated in advance of the predicted onset week

(Table 1). Specifically, no temperate forecasts predicted that onset would occur with more than a one week advanced lead, and very few forecasts in the tropics accurately predicted onset with more than a one-week lead. In temperate regions, onset timing accuracy (onset predicted within one week of the observed value) quickly increased and remained above 95% as soon as the predicted onset was in the past. In the tropics, accuracy reached almost 50% at the predicted onset, and remained around 65-70% for all later lead weeks.

For the tropics only, we also evaluated how often forecasts correctly recognized an existing or upcoming outbreak, without mistakenly predicting outbreaks during periods in which no outbreaks occurred. Specifically, we calculated sensitivity, specificity, positive predictive value, and negative predictive value. We found that both sensitivity (98.56%) and the negative predictive value (98.10%) were high, but that specificity (56.22%) and the positive predictive value (63.12%) were much lower. Thus, while forecasts are unlikely to predict dormancy before or during an outbreak, forecasts suggesting a current or upcoming outbreak were often inaccurate.

Comparison to Method of Analogues

We also compared our forecasting results to results obtained using the method of analogues, a non-mechanistic forecasting method previously used by Viboud et al. (2003) to forecast influenza incidence in France. In temperate countries, our mechanistic forecasting approach outperformed the method of analogues slightly for peak timing, and substantially for peak intensity before the predicted peak (S13 Figure A and B). In the tropics, the two methods performed similarly for both peak timing and intensity (S13 Figure C and D) before the peak, and the method of analogues performed slightly better after the predicted peak. Thus, the

mechanistic forecasting methods used in this work only improve upon the analogue forecasting method in temperate regions. Additional details can be found in S1 Text, and in Viboud et al. (2003).

Factors Influencing Forecast Accuracy

Temperate vs. Tropical Regions

As expected, forecast accuracy was significantly lower in the tropics than in temperate regions. Overall, the odds that a forecast accurately predicted peak timing in the tropics was 0.123 (95% CI: 0.091, 0.165) times that in temperate regions, and the odds of accurately predicting peak intensity in the tropics were 0.103 (95% CI: 0.072, 0.148) times that in temperate regions. This pattern held when comparisons were restricted to predicted lead weeks of 0 and greater (i.e. forecasts predicting that the peak was either the current week or in the past; peak timing aOR = 0.115 (0.084, 0.160); peak intensity aOR = 0.070 (0.045, 0.108)).

Impact of Humidity Forcing

Inclusion of humidity forcing in the temperate region forecasts significantly increased both peak timing and peak intensity forecast accuracy prior to the observed peak, and peak timing accuracy at the observed peak (Table 2). Post-peak, no significant differences in forecast accuracy were observed by inclusion of humidity forcing.

Additional Factors

Forecast accuracy was also assessed by hemisphere, region, data type, season, and scaling in the temperate regions, and by region, data type, and scaling in the tropics. Few consistent,

significant relationships were found. In temperate regions, peak timing accuracy was lower for countries reporting ARI+ data vs. ILI+ data both before (aOR = 0.645, 95% CI: 0.428-0.973) and after (aOR = 0.567, 95% CI: 0.334-0.965) the predicted peak. Peak timing accuracy was highest after the peak in Eastern Europe (aOR = 2.068, 95% CI: 1.095-3.889, compared to Southwest Europe; see S1 Text for information on how countries were classified into regions). Finally, compared to countries with scaling values between 2 and 10, countries using scaling values between 0 and 0.5 performed worse for both peak timing (aOR = 0.420, 95% CI: 0.181-0.982) and peak intensity (aOR = 0.170, 95% CI: 0.051-0.568) after the predicted peak. Post-peak, countries using scaling values between 10 and 20 (aOR = 0.145, 95% CI: 0.038-0.571), 20 and 100 (aOR = 0.147, 95% CI: 0.033-0.646), and 100 and 500 (aOR = 0.229, 95% CI: 0.066-0.801) also performed significantly worse for peak intensity only. No significant differences in forecast accuracy were observed by hemisphere or season for either peak timing or intensity (results in S3 Table).

Table 2. Accuracy of forecasts incorporating vs. omitting absolute humidity forcing by observed lead week for both peak timing and intensity.

Obs. Lead Week:		-6	-5	-4	-3	-2	-1	0	1	2	3	4
Timing	AH	4.8%	18.6%	36.7%	47.3%	54.4%	53.6%	55.4%	70.9%	80.2%	82.1%	82.0%
	No AH	2.0%	13.2%	32.3%	41.8%	50.4%	48.3%	50.4%	71.6%	81.4%	83.6%	83.3%
	Sig.	**	**	**	**	**	**	**				
Intensity	AH	6.9%	10.1%	18.9%	26.3%	40.3%	56.8%	70.1%	86.7%	90.2%	91.0%	90.1%
	No AH	4.2%	7.5%	14.1%	25.2%	41.6%	52.8%	68.4%	87.4%	90.4%	91.8%	91.1%
	Sig.	**	**	**			*					

* p<0.001, **p<0.0001

Because very few forecasts were generated prior to the predicted peak week in the tropics, it was only possible to rigorously compare forecast accuracy at and after the predicted peak. No statistically significant associations between forecast accuracy and data type, region, or scaling value were found for either peak timing or intensity in the tropics (results in S4 Table).

Forecast Calibration

It is important to consider not only how accurate forecasts are, but also forecast uncertainty. This is especially true in the case of real-time forecasting: different medical and public health responses might be affected given forecast of an 80% chance of a particular outcome rather than a 20% chance. Because each forecast is based on 300 individual ensemble members, we could assess forecast certainty through the spread of the ensemble variance, where narrower ensemble spread ideally indicated greater certainty.

Figures 3A and B show average peak timing and intensity forecast accuracy, respectively, for temperate regions plotted against ensemble variance (separated into 10 quantiles). For peak timing, we generally saw a slight decrease in forecast accuracy as ensemble variance increases at all predicted lead weeks, indicating that we can infer expected forecast accuracy from ensemble spread. For peak intensity, this pattern only held prior to the predicted peak. Corresponding plots for the tropics are shown in Figures 3C and D. For peak timing, no clear relationship existed between ensemble variance and forecast accuracy, indicating that no information about expected forecast accuracy can be inferred from ensemble spread. For forecasts of peak intensity, on the other hand, increases in ensemble variance corresponded to substantial decreases in forecast accuracy.

We also explored how often the observed peak timing and intensity fall within certain prediction intervals of ensemble spread prior to the predicted peak (Figure 4). In a well-calibrated forecast, we expect that the observed intensity will fall within the *n*th% prediction interval *n*% of the time. Overall, forecasts appeared to be well calibrated for both peak timing and intensity in temperate regions at all lead times, although prediction intervals tended to be too wide for peak timing, especially several weeks before the peak. In the tropics, peak intensity

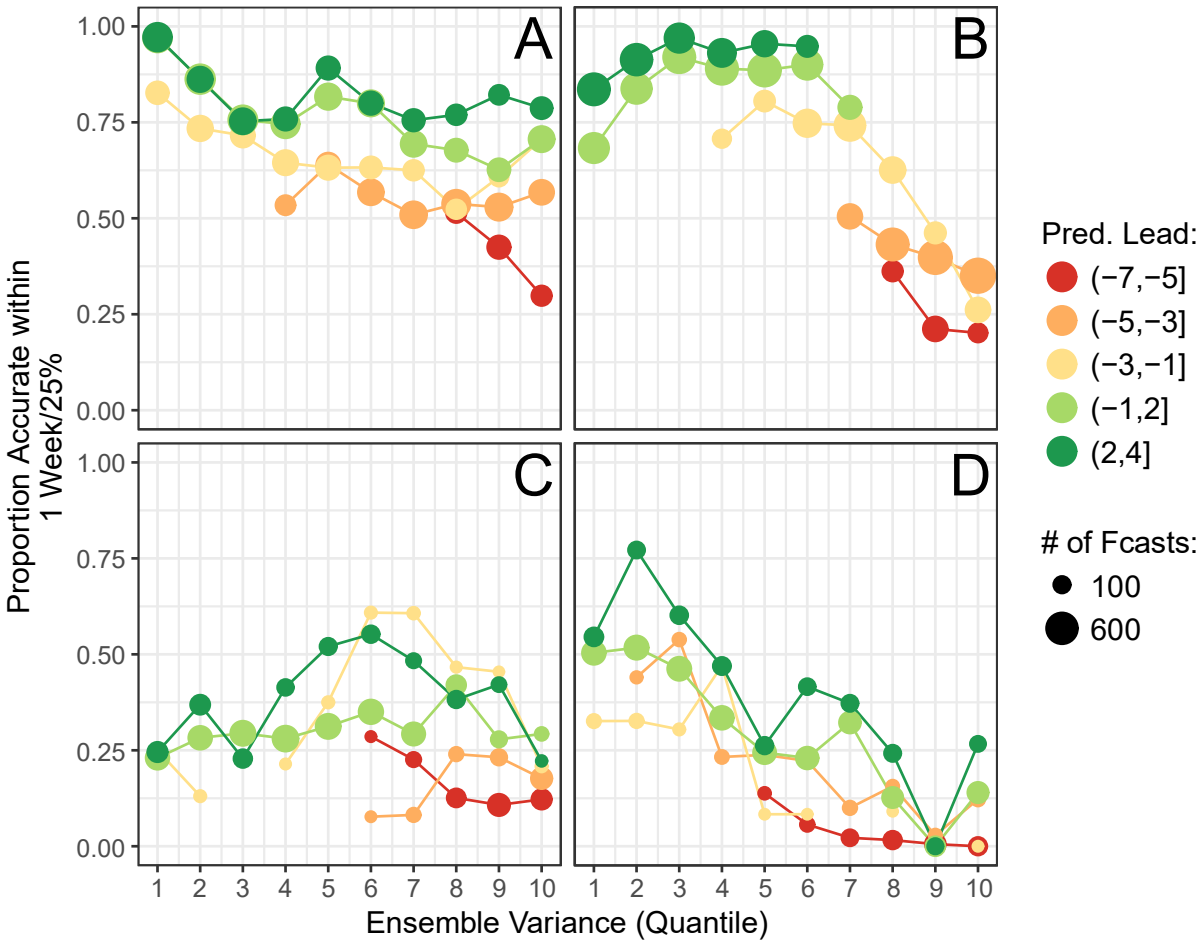


Figure 3. Forecast calibration as the relationship between forecast accuracy and ensemble spread. The relationship between peak timing (A and C) and peak intensity (B and D) ensemble variance and forecast accuracy by predicted lead week is shown for temperate (A and B) and tropical (C and D) regions. Point size represents how many forecasts are included in the point, and only lead week ranges with at least 100 (A and B) or 10 (C and D) forecasts were included.

forecasts appeared well calibrated, while peak timing forecasts rarely included the observed peak timing.

Further exploration of forecast calibration can be found in S10 Figure.

Discussion

While skillful forecasts of influenza activity have repeatedly been shown to be possible (Biggerstaff et al., 2016; Brooks et al., 2015; Hickmann et al., 2015; Morita et al., 2018; Moss et

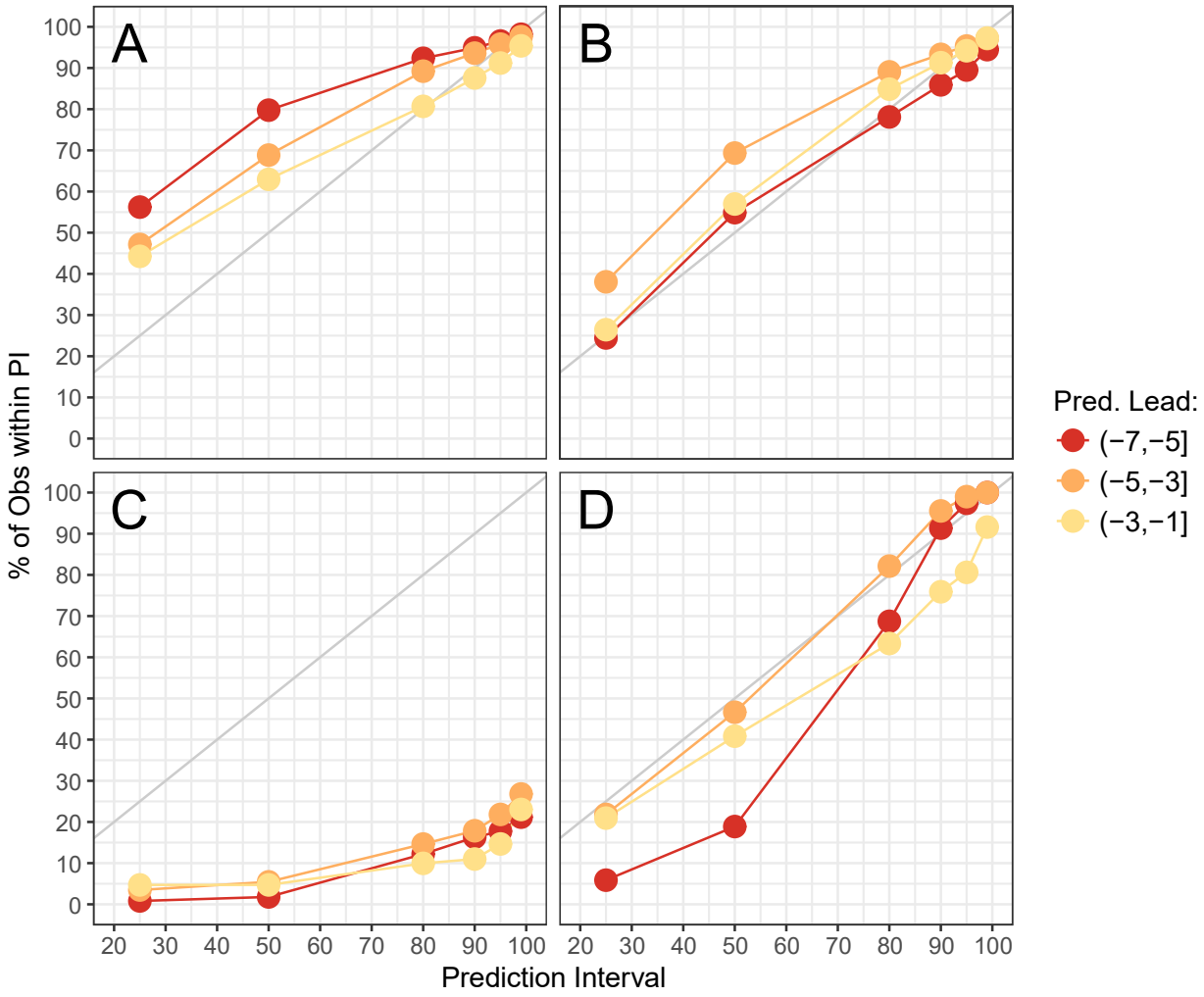


Figure 4. Forecast calibration as percent of observed peak timing/intensity values falling within prediction intervals.
Here we show the percentage of forecasts where the observed peak timing (A and C) or intensity (B and D) value falls within the 25%, 50%, 80%, 90%, 95%, and 99% prediction intervals of 300 ensemble members by predicted lead week for temperate (A and B) and tropical (C and D) regions. The gray line represents the expected case in which exactly 25% of observations are falling within the 25% prediction interval, and so on.

al., 2016, 2017; Ong et al., 2010; Osthus et al., 2017; Shaman et al., 2013; Shaman & Karspeck, 2012; Viboud et al., 2003), few attempts to forecast non-pandemic influenza outbreaks in areas other than the US have been made. Here, we use publicly-available syndromic and virologic data to generate retrospective forecasts of influenza transmission at the country scale for 64 countries in both temperate and tropical regions. We find that accurate and well-calibrated forecasts are possible in temperate regions. On average, peak timing and peak intensity of outbreaks can be

predicted within 1 week and within 25% of the observed values, respectively, over 50% of the time starting four (peak timing) and two (peak intensity) weeks before the predicted peak, although forecast accuracy differs substantially by country (results in S2 Figure). These results are broadly consistent with past forecasting results in various US cities (Shaman et al., 2013; Shaman & Karspeck, 2012) as well as in Victoria, Australia (Moss et al., 2016, 2017), indicating that the larger spatial scale employed here does not substantially compromise forecast accuracy.

As expected, forecasts were both less accurate and less well calibrated in the tropics. Typically, peak timing and intensity could not be predicted within 1 week or 25% of observed values, respectively, until after the peak was estimated to have occurred, and the proportion of forecasts achieving these accuracy levels never exceeded 50%, even several weeks after the peak. For peak timing in particular, prediction intervals based on 300 ensemble runs rarely included the observed peak week, and ensemble variability was not strongly associated with forecast accuracy, making forecast calibration challenging. Finally, while sensitivity and the negative predictive value were high, specificity and the positive predictive value were low, indicating that forecasts often predicted outbreaks when no outbreaks occurred in reality. Previously, Yang et al. (2015) produced forecasts of non-pandemic influenza in Hong Kong using methods similar to those employed here, and found that both peak timing and intensity accuracy reached 50% by lead week 0. On average, our tropical forecasts perform more poorly, although we note that, as in temperate regions, forecast performance varied substantially by country (results in S2 Figure). It is possible that the data for many of the tropical countries used in forecasting here are simply noisier than the Hong Kong data. If so, this issue may be difficult to surmount without changes in surveillance methods: smoothing our tropical data using a simple moving average over three weeks did not substantially improve forecast accuracy (results in S1

Text and S8 Figure), nor did performing model fitting and forecasting by individual outbreak instead of continuously across multiple outbreaks (results in S1 Text and S9 Figure). We also emphasize that, while our method is well tested in temperate regions, very little forecasting has been performed in the tropics. Our results do not suggest tropical countries will always yield forecasts with low accuracy, simply that the combination of data and methods applied in temperate countries may be insufficient.

Unlike peak timing and intensity, onset timing was not accurately predicted before outbreak onset in either temperate or tropical regions. This poor performance is likely due to a lack of signal in the data prior to the start of an outbreak, and is not surprising. Past work has shown that models including travel between US states (Pei et al., 2018) and boroughs of New York City (Yang et al., 2016) significantly improve forecast accuracy, particularly onset timing accuracy. Future work will incorporate travel between countries in the model, allowing forecasts of onset timing in a given country to be informed by signal from connected countries in which an outbreak has already begun. While a variety of models exist for forecasting the spatial dynamics of influenza transmission (Geilhufe et al., 2014; Longini et al., 1986; Tizzoni et al., 2012), we believe that our approach, in which a model is iteratively fit to influenza observations, can offer significant improvements.

Significant differences in forecast accuracy were observed by a variety of factors for both peak timing and intensity. In temperate regions, forecasts of peak timing are less accurate for countries reporting ARI data than for those reporting ILI data. Because ARI is a less specific measure than ILI, these data tend to be noisier. This, in turn, likely contributes to the lower forecast accuracy. We also observe lower peak timing accuracy with particularly small scaling

values, and lower peak intensity accuracy with particularly small or large scaling values, at least after the peak.

Including absolute humidity forcing in our models improved temperate forecast accuracy prior to the observed peak, with no differences observed post-peak. These results are consistent with the results of a recent paper by Shaman et al. (2017), which found that, on average, including absolute humidity forcing improved forecast accuracy in 95 US cities before the predicted peak. Given the large spatial and latitudinal scale of several of the countries examined here, it is interesting that mean country-level absolute humidity still significantly improves forecast accuracy. Our results suggest that absolute humidity forcing should continue to be included in models forecasting influenza in temperate regions, even when humidity must be averaged over large regions. Furthermore, given evidence that climatic factors such as absolute and relative humidity (Emukule et al., 2016; Imai et al., 2014; Soebiyanto et al., 2014; Tamerius et al., 2013) and precipitation (Imai et al., 2014; Soebiyanto et al., 2014; Tamerius et al., 2013; Yang et al., 2018) may influence influenza transmission in the tropics and subtropics, future work should consider how climatic factors may be incorporated into model fitting and forecasting outside of the temperate zones.

Despite the novelty of this work, we are cognizant of several important limitations. First, our data exhibit a strong spatial bias, with little to no representation in Africa and South America. Information on influenza dynamics in general are particularly lacking in the tropics, which precludes forecasting. As always, forecast accuracy findings may also be dependent on the choice of accuracy metrics, although we note that our results are robust to various accuracy cutoffs, as well as to choice of alternative accuracy metrics (see S1 Text; S11 and S12 Figures).

Furthermore, our method outperforms the method of analogues, a robust non-mechanistic forecasting method (Viboud et al., 2003) in temperate regions (S1 Text and S13 Figure).

Additionally, as mentioned in the Methods, data do not perfectly reflect reality. All syndromic data types include some people with non-influenza respiratory conditions, and exclude those with influenza who do not seek treatment or meet specific criteria. Multiplying by percent of tests positive for influenza only partially mitigates these issues. In particular, differences in noisiness between different data types persist. Information on the size of the catchment area from which data were obtained is also largely lacking, so forecasts must be generated based on raw counts rather than rates. This leads to substantial variability in case counts by country. While this can be partially compensated through the use of scaling factors, it is crucial that these values are chosen appropriately (Moss et al., 2017). As we base scaling values on past data, forecast accuracy may therefore be compromised when data are not available for several past seasons. Furthermore, if new countries begin submitting influenza data, real-time forecasts cannot be generated immediately, as at least one full season or outbreak must pass before an appropriate scaling can be calculated.

Finally, all forecasts at this point have been generated at the country level. Thus, while our results and future real-time forecasts may be of public health relevance for smaller countries, they are likely to provide less actionable results for much larger countries, such as Russia or Brazil. Future work should attempt to incorporate subnational data, where available. In addition to increased public health relevance, we may also expect forecast accuracy to improve when smaller subunits within a country are used for forecasting. We note, however, that real-time forecasts are only plausible when data are submitted in a timely fashion. At present, this occurs

for most of the northern hemisphere temperate countries included in this study, but is uncommon in southern hemisphere temperate countries and for countries in the tropics and subtropics.

Conclusions

We have shown that, in temperate regions, accurate and well-calibrated retrospective forecasts of seasonal influenza activity are feasible. Work is currently being conducted to determine whether real-time forecasts are similarly feasible, and future work will incorporate travel between countries with the goal of improving forecast accuracy, particularly onset timing accuracy. Although this work is at an early stage, we note the importance of eventually incorporating forecasts into medical and public health decision-making. Accurate real-time probabilistic forecasts have the potential to inform decisions such as antiviral stockpiling by governments or staff and bed management by hospitals, preventing morbidity and mortality. Therefore, it is critical that these forecasts not be produced solely as an academic exercise.

References

Adde, A., Roucou, P., Mangeas, M., Ardillon, V., Desenclos, J.-C., Rousset, D., Girod, R., Briolant, S., Quenel, P., & Flamand, C. (2016). Predicting Dengue Fever Outbreaks in French Guiana Using Climate Indicators. *PLOS Neglected Tropical Diseases*, *10*(4), e0004681. https://doi.org/10.1371/journal.pntd.0004681

Anderson, J. L. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;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., Velardi, P., Vespignani, A., & Finelli, L. (2016). Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infectious Diseases*, *16*. https://doi.org/10.1186/s12879-016-1669-x

Bloom-Feshbach, K., Alonso, W. J., Charu, V., Tamerius, J., Simonsen, L., Miller, M. A., & Viboud, C. (2013). Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PLOS ONE*, *8*(2), e54445. https://doi.org/10.1371/journal.pone.0054445

Brooks, L. C., Farrow, D. C., Hyun, S., Tibshirani, R. J., & Rosenfeld, R. (2015). Flexible Modeling of Epidemics with an Empirical Bayes Framework. *PLoS Computational Biology*, *11*(8). https://doi.org/10.1371/journal.pcbi.1004382

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M. A., Flasche, S., Baguelin, M., Pollington, T., Carney, J. R., Glover, R., Smout, E., Tiffany, A., Edmunds, W. J., & Funk, S. (2015). Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: A Real-time Modelling Study. *PLOS Currents Outbreaks*. https://doi.org/10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2

Chen, I.-F., & Lu, C.-J. (2017). Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications*, *28*(9), 2633–2647. https://doi.org/10.1007/s00521-016-2215-x

Choi, T.-M., Hui, C.-L., Liu, N., Ng, S.-F., & Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, *59*, 84–92. https://doi.org/10.1016/j.dss.2013.10.008

Chowell, G., Hincapie-Palacio, D., Ospina, J., Pell, B., Tariq, A., Dahal, S., Moghadas, S., Smirnova, A., Simonsen, L., & Viboud, C. (2016). Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics. *PLoS Currents*, *8*. https://doi.org/10.1371/currents.outbreaks.f14b2217c902f453d9320a43a35b9583

Debry, E., & Mallet, V. (2014). Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and PM10 on the Prev'Air platform. *Atmospheric Environment*, *91*, 71–84. https://doi.org/10.1016/j.atmosenv.2014.03.049

Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., & Sugihara, G. (2016). Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences*, *113*(46), 13081–13086. https://doi.org/10.1073/pnas.1607747113

Emukule, G. O., Mott, J. A., Spreeuwenberg, P., Viboud, C., Commanday, A., Muthoka, P., Munywoki, P. K., Nokes, D. J., van der Velden, K., & Paget, J. W. (2016). Influenza activity in Kenya, 2007-2013: Timing, association with climatic factors, and implications for vaccination campaigns. *Influenza and Other Respiratory Viruses*, *10*(5), 375–385. https://doi.org/10.1111/irv.12393

FAO. (2016). *Crop yield forecasting: Methodological and institutional aspects.* Food and Agriculture Organization of the United Nations. http://gsars.org/en/crop-yield-forecasting-methodological-and-institutional-aspects/

Flu News Europe | System. (n.d.). Retrieved May 17, 2018, from https://flunewseurope.org/System

Funk, S., Camacho, A., Kucharski, A. J., Eggo, R. M., & Edmunds, W. J. (2018). Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics*, *22*, 56–61. https://doi.org/10.1016/j.epidem.2016.11.003

Gaubert, B., Coman, A., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., & Beekmann, M. (2014). Regional scale ozone data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model. *Geosci. Model Dev.*, *7*(1), 283–302. https://doi.org/10.5194/gmd-7-283-2014

Geilhufe, M., Held, L., Skrøvseth, S. O., Simonsen, G. S., & Godtliebsen, F. (2014). Power law approximations of movement network data for modeling infectious disease spread: Power law approximations of movement network data. *Biometrical Journal*, *56*(3), 363–382. https://doi.org/10.1002/bimj.201200262

Gneiting, T., & Raftery, A. E. (2005). Atmospheric science. Weather forecasting with ensemble methods. *Science (New York, N.Y.)*, *310*(5746), 248–249. https://doi.org/10.1126/science.1115255

Held, L., Meyer, S., & Bracher, J. (2017). Probabilistic forecasting in infectious disease epidemiology: The 13th Armitage lecture. *Statistics in Medicine*, *36*(22), 3443–3460. https://doi.org/10.1002/sim.7363

Hickmann, K. S., Fairchild, G., Priedhorsky, R., Generous, N., Hyman, J. M., Deshpande, A., & Del Valle, S. Y. (2015). Forecasting the 2013–2014 Influenza Season Using Wikipedia. *PLoS Computational Biology*, *11*(5). https://doi.org/10.1371/journal.pcbi.1004239

Huff, A., Allen, T., Whiting, K., Breit, N., & Arnold, B. (2016). FLIRT-ing with Zika: A Web Application to Predict the Movement of Infected Travelers Validated Against the Current Zika Virus Epidemic. *PLOS Currents Outbreaks*. https://doi.org/10.1371/currents.outbreaks.711379ace737b7c04c89765342a9a8c9

Imai, C., Brooks, W. A., Chung, Y., Goswami, D., Anjali, B. A., Dewan, A., Kim, H., & Hashizume, M. (2014). Tropical influenza and weather variability among children in an urban low-income population in Bangladesh. *Global Health Action*, *7*(1), 24413. https://doi.org/10.3402/gha.v7.24413

Johnson, L. R., Gramacy, R. B., Cohen, J., Mordecai, E., Murdock, C., Rohr, J., Ryan, S. J., Stewart-Ibarra, A. M., & Weikel, D. (2017). Phenomenological forecasting of disease incidence using heteroskedastic Gaussian processes: A dengue case study. *ArXiv:1702.00261 [q-Bio, Stat]*. http://arxiv.org/abs/1702.00261

Kandula, S., Yang, W., & Shaman, J. (2017). Type- and Subtype-Specific Influenza Forecast. *American Journal of Epidemiology*, *185*(5), 395–402. https://doi.org/10.1093/aje/kww211

Longini, I. M., Fine, P. E. M., & Thacker, S. B. (1986). PREDICTING THE GLOBAL SPREAD OF NEW INFECTIOUS AGENTS. *American Journal of Epidemiology*, *123*(3), 383–391. https://doi.org/10.1093/oxfordjournals.aje.a114253

Mccarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, *25*(5), 303–324. https://doi.org/10.1002/for.989

Meltzer, M. I., Atkins, C. Y., Santibanez, S., Knust, B., Petersen, B. W., Ervin, E. D., Nichol, S. T., Damon, I. K., Washington, M. L., & Centers for Disease Control and Prevention (CDC). (2014). Estimating the future number of cases in the Ebola epidemic—Liberia and Sierra Leone, 2014-2015. *MMWR Supplements*, *63*(3), 1–14.

Morita, H., Kramer, S., Heaney, A., Gil, H., & Shaman, J. (2018). Influenza forecast optimization when using different surveillance data types and geographic scale. *Influenza and Other Respiratory Viruses*. https://doi.org/10.1111/irv.12594

Moss, R., Zarebski, A., Dawson, P., & McCaw, J. M. (2016). Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza and Other Respiratory Viruses*, *10*(4), 314–323. https://doi.org/10.1111/irv.12376

Moss, R., Zarebski, A., Dawson, P., & McCAW, J. M. (2017). Retrospective forecasting of the 2010–2014 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology & Infection*, *145*(1), 156–169. https://doi.org/10.1017/S0950268816002053

Newlands, N. K., Zamar, D. S., Kouadio, L. A., Zhang, Y., Chipanshi, A., Potgieter, A., Toure, S., & Hill, H. S. J. (2014). An integrated, probabilistic model for improved seasonal

forecasting of agricultural crop yield under environmental uncertainty. *Frontiers in Environmental Science*, *2*. https://doi.org/10.3389/fenvs.2014.00017

Ng, S., & Gordon, A. (2015). Influenza Burden and Transmission in the Tropics. *Current Epidemiology Reports*, *2*(2), 89–100. https://doi.org/10.1007/s40471-015-0038-4

Nishiura, H. (2011). Real-time forecasting of an epidemic using a discrete time stochastic model: A case study of pandemic influenza (H1N1-2009). *BioMedical Engineering OnLine*, *10*, 15. https://doi.org/10.1186/1475-925X-10-15

Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., & Marathe, M. V. (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses*, *8*(3), 309–316. https://doi.org/10.1111/irv.12226

Ong, J. B. S., Chen, M. I.-C., Cook, A. R., Lee, H. C., Lee, V. J., Lin, R. T. P., Tambyah, P. A., & Goh, L. G. (2010). Real-Time Epidemic Monitoring and Forecasting of H1N1-2009 Using Influenza-Like Illness from General Practice and Family Doctor Clinics in Singapore. *PLoS ONE*, *5*(4), e10036. https://doi.org/10.1371/journal.pone.0010036

Osthus, D., Hickmann, K. S., Caragea, P. C., Higdon, D., & Del Valle, S. Y. (2017). Forecasting seasonal influenza with a state-space SIR model. *The Annals of Applied Statistics*, *11*(1), 202–224. https://doi.org/10.1214/16-AOAS1000

Pei, S., Kandula, S., Yang, W., & Shaman, J. (2018). Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences*, 201708856. https://doi.org/10.1073/pnas.1708856115

Reich, N. G., Lauer, S. A., Sakrejda, K., Iamsirithaworn, S., Hinjoy, S., Suangtho, P., Suthachana, S., Clapham, H. E., Salje, H., Cummings, D. A. T., & Lessler, J. (2016). Challenges in Real-Time Prediction of Infectious Disease: A Case Study of Dengue in Thailand. *PLoS Neglected Tropical Diseases*, *10*(6), e0004761. https://doi.org/10.1371/journal.pntd.0004761

Rodell, M. (n.d.). *LDAS | Land Data Assimilation Systems* [Text.Journal]. https://ldas.gsfc.nasa.gov/gldas/GLDASgoals.php

Shaman, J., Kandula, S., Yang, W., & Karspeck, A. (2017). The use of ambient humidity conditions to improve influenza forecast. *PLOS Computational Biology*, *13*(11), e1005844. https://doi.org/10.1371/journal.pcbi.1005844

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, *109*(50), 20425–20430. https://doi.org/10.1073/pnas.1208772109

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3837

Shaman, J., & Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3243–3248. https://doi.org/10.1073/pnas.0806852106

Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T., & Lipsitch, M. (2010). Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States. *PLoS Biology*, *8*(2), e1000316. https://doi.org/10.1371/journal.pbio.1000316

Shaman, J., Yang, W., & Kandula, S. (2014). Inference and forecast of the current west african ebola outbreak in Guinea, sierra leone and liberia. *PLoS Currents*, *6*. https://doi.org/10.1371/currents.outbreaks.3408774290b1a0f2dd7cae877c8b8ff6

Shi, Y., Liu, X., Kok, S.-Y., Rajarethinam, J., Liang, S., Yap, G., Chong, C.-S., Lee, K.-S., Tan, S. S. Y., Chin, C. K. Y., Lo, A., Kong, W., Ng, L. C., & Cook, A. R. (2015). Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environmental Health Perspectives*, *124*(9). https://doi.org/10.1289/ehp.1509981

Soebiyanto, R. P., Clara, W., Jara, J., Castillo, L., Sorto, O. R., Marinero, S., de Antinori, M. E. B., McCracken, J. P., Widdowson, M.-A., Azziz-Baumgartner, E., & Kiang, R. K. (2014). The Role of Temperature and Humidity on Seasonal Influenza in Tropical Areas: Guatemala, El Salvador and Panama, 2008–2013. *PLoS ONE*, *9*(6), e100659. https://doi.org/10.1371/journal.pone.0100659

Tamerius, J. D., Shaman, J., Alonso, W. J., Bloom-Feshbach, K., Uejio, C. K., Comrie, A., & Viboud, C. (2013). Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLoS Pathogens*, *9*(3), e1003194. https://doi.org/10.1371/journal.ppat.1003194

Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J. J., Balcan, D., Gonçalves, B., Perra, N., Colizza, V., & Vespignani, A. (2012). Real-time numerical forecast of global epidemic spreading: Case study of 2009 A/H1N1pdm. *BMC Medicine*, *10*(1), 165.

Viboud, C., Boëlle, P.-Y., Carrat, F., Valleron, A.-J., & Flahault, A. (2003). Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, *158*(10), 996–1006.

WHO. (n.d.-a). *Influenza (Seasonal)*. WHO. https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)

WHO. (n.d.-b). *WHO | FluID - a global influenza epidemiological data sharing platform*. WHO. https://www.who.int/influenza/surveillance_monitoring/fluid/en/

WHO. (n.d.-c). *WHO | FluNet*. WHO. https://www.who.int/influenza/gisrs_laboratory/flunet/en/

Yang, W., Cowling, B. J., Lau, E. H. Y., & Shaman, J. (2015). Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology*, *11*(7), e1004383. https://doi.org/10.1371/journal.pcbi.1004383

Yang, W., Cummings, M. J., Bakamutumaho, B., Kayiwa, J., Owor, N., Namagambo, B., Byaruhanga, T., Lutwama, J. J., O'Donnell, M. R., & Shaman, J. (2018). Dynamics of influenza in tropical Africa: Temperature, humidity, and co-circulating (sub)types. *Influenza and Other Respiratory Viruses*, *12*(4), 446–456. https://doi.org/10.1111/irv.12556

Yang, W., Karspeck, A., & Shaman, J. (2014). Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLoS Computational Biology*, *10*(4), e1003583. https://doi.org/10.1371/journal.pcbi.1003583

Yang, W., Olson, D. R., & Shaman, J. (2016). Forecasting Influenza Outbreaks in Boroughs and Neighborhoods of New York City. *PLoS Computational Biology*, *12*(11), e1005201.

Zebiak, S. E., & Cane, M. (1987). A model El Niño-Southern Oscillation. *Monthly Weather Review*, *115*, 2262–2278.

Zebiak, S. E., Orlove, B., Muñoz, Á. G., Vaughan, C., Hansen, J., Troy, T., Thomson, M. C., Lustig, A., & Garvin, S. (2015). Investigating El Niño-Southern Oscillation and society relationships. *Wiley Interdisciplinary Reviews: Climate Change*, *6*(1), 17–34. https://doi.org/10.1002/wcc.294

S1 Text: Supplementary methods and results.

Supplementary Methods

Influenza Data Processing

Both syndromic and virologic data were downloaded from the WHO in April 2018.

When a country reported several different types of syndromic data, we used the data type that was most consistently reported across all seasons. When two data types were reported with roughly equal frequency, we favored ILI, as it is more specific than ARI, but includes a wider population than SARI or pneumonia. Data were also examined visually for signal; if one data type appeared to produce a much smoother signal than another, it was chosen for use in forecasting. France reported ARI data prior to the 2014-15 influenza season, but switched to ILI data for the 2014-15 season and all subsequent seasons; all other countries have favored the same data type or types over time.

Overall, 38 countries reported ILI, 17 reported ARI, 6 reported SARI, one (Honduras) reported pneumonia, one (Canada) reported ILI rates rather than counts, and one (France) changed preferential data types during the period spanned by the data (from ARI to ILI) (S1 Table). SARI and pneumonia were only preferentially reported from tropical countries. Definitions for all 4 syndromes, however, are not standardized, and therefore the specific definitions used vary by member state. Broadly, ILI refers to a respiratory illness involving fever and cough, whereas ARI is less strict and captures patients with at least one of several respiratory symptoms. A diagnosis of SARI, meanwhile, requires hospitalization (WHO, 2011).

Virologic data consisted of the number of tests positive for any influenza strain, as well as the number of tests processed and reported. The proportion of tests positive for influenza was

calculated by dividing the number of positive tests by the number of tests processed; when no information regarding tests processed was available, the number of positive tests was instead divided by the number of tests reported. If the resulting proportion exceeded one, the data point was removed.

Countries were maintained in the dataset if they had good quality syndromic and virologic data. In the temperate regions, good quality data were defined as data for which fewer than one third of all available seasons met the removal criteria described in the main text (i.e. at least 5 consecutive missing data points near the peak); in other words, temperate countries were maintained in the dataset if over two thirds of available seasons could be used for forecasting. In the tropics, countries were removed from consideration if: 1) over 33% of total observations or over 5% of observations during outbreaks were 0, or 2) if the highest peak was over 15 times higher than the lowest peak (i.e., if the data showed an unrealistic amount of variation from outbreak to outbreak).

Finally, several countries were located between temperate and tropical regions in the subtropics, whereas others spanned both temperate and tropical regions. To classify these countries as “temperate” or “tropical” for the sake of this study, we therefore considered whether past influenza outbreaks exhibited a marked seasonal signal consistent with temperate influenza activity. If outbreaks occurred once a year and strictly within the seasons defined for temperate regions (weeks 40 to 20 in the northern hemisphere, or weeks 14 to 46 in the southern), the country was classified as “temperate;” otherwise, we classified it as “tropical.”

All data used for forecasting are described in S1 and S2 Tables, and visualized in S1 Figure.

Humidity Data Processing

We processed the raw data for use in our models as follows. First, daily averages for each 1°x1° grid cell were calculated to yield daily time series of specific humidity. Next, we searched the humidity data visually for anomalies. Three major anomalies were found: 1) in some grid cells, humidity increased substantially for the years 1994-1998; 2) in some grid cells, humidity was anomalously low in either 1999 or 2003-2004; and 3) in several countries, primarily Australia and countries in Eastern Europe, a sharp increase in humidity was observed throughout the majority of 1997, excepting the summer. The first two anomalies were addressed by removing any years for which yearly average specific humidity was over 1.5 times the 75th percentile or less than 0.65 times the 25th percentile of yearly average specific humidity for a given grid cell. The third anomaly was identified by visual inspection, and the year 1997 was removed from affected grid cells. A total of 469 grid cells were affected by the first anomaly, 44 by the second, and 1270 by the third; additionally, two grid cells from Chile were removed entirely because data shifted substantially upward and downward over time. Overall, 1800 grid cells had at least some data removed due to anomalies, leaving 6240 grid cells with no anomalies during the 20-year record.

Twenty-year climatologies for each grid cell were then generated by averaging daily specific humidity on each of 365 days across twenty years. Note that, due to removal of anomalous data by year, many grid cells yielded 12 to 19 year climatologies. Each grid cells was then assigned to one or more countries using the Clip tool in QGIS 2.18.2. Grid cells belonging to more than one country were delegated proportionally to all countries with which they overlapped. Finally, climatologies were aggregated to the country level by taking an average of

the climatologies for all grid cells assigned to a given country, weighted by the proportion of the grid cell situated within the country in question.

Ensemble Adjustment Kalman Filter

Observational Error Variance: As described in the main text and above, both syndromic and virologic data are subject to error from a variety of sources, and thus deviate from the number of true influenza cases. However, the extent of this error is unknown. In order to properly use the EAKF, we must therefore specify some degree of error in our observations. We account for this by calculating observational error variance (OEV), defined as:

$$OEV_t = \frac{1 \times 10^5 + \frac{\left(\sum_{j=t-3}^{t-1} \frac{O_t}{3}\right)^2}{5}}{c}$$

where O_t is the observed syndromic+ data at time t and c can be altered to modify the magnitude of the prescribed error, with lower values of c corresponding to higher overall error in the observations. All forecasts described in the main text were run with c equal to 1. Results of sensitivity analyses using c = 10 are presented below.

Filter Divergence: One prominent issue with the EAKF is that of filter divergence, in which, following assimilation of multiple successive observations, the variance of the model ensemble decreases, and thus confidence in the model estimates increases to the point where the observations are essentially ignored. To prevent filter divergence in temperate regions, we multiplicatively inflate the prior model variance by 1.03 times before assimilating each new observation (Shaman et al., 2013; Yang, Cowling, et al., 2015). In the tropics, where model

fitting is performed over several years, filter divergence is likely to be a more substantial issue than over the shorter, seasonal time periods modeled for temperate countries. As in the temperate regions, we address filter divergence by multiplicatively inflating the ensemble variance by 1.03 at each time step. Additionally, per Yang, Cowling et al. (2015), if the model diverges from an observation by more than 20% at a given time step, we reinitialize the model completely by choosing new initial states and parameters at that time step. These methods are described in more detail in Yang, Cowling et al. (2015).

Retrospective Forecast Generation (Temperate without Humidity Forcing)

To generate retrospective forecasts in temperate regions with no absolute humidity forcing, we used the same SIRS model as for tropical forecasts. As with all other forecasts, we ran 5 simulations of 300 ensemble members each.

Comparing Forecast Accuracy

To compare forecast accuracy in the temperate and tropical regions, as well as by hemisphere, region, season, data type, and chosen scaling value, we used generalized estimating equations (GEEs) controlling for predicted lead week as a categorical variable with week 0 as the reference level. GEEs were chosen for their ability to control for temporal autocorrelation within each country and season pair, as the accuracy of successive weekly forecasts in a given country are temporally autocorrelated. Further, GEEs were chosen over mixed effects models in order to estimate overall effects rather than impact on individual forecasts. An autoregressive AR(1) working correlation matrix was assumed. The five replicate forecasts produced for each country, season, and start week represent an additional layer of clustering in our results. To control for

this, we randomly permuted the results 100 times, each time choosing a single run (among the 5 replicates) for each country, season, and start week (or, in the tropics, each country, start week, and individual outbreak). Final results were drawn from the median coefficients and standard errors of these 100 permutations. Results for all tested factors for both temperate and tropical regions can be found in S3 and S4 Tables.

Tropical Data Smoothing

We hypothesized that forecast accuracy in the tropics could be improved by smoothing the syndromic+ data, which was typically substantially noisier than the temperate data (S1 Figure). In order to test this hypothesis, we applied a simple moving average to the tropical data. Because in real-time forecasting no data beyond the current forecast week are available, we averaged the data for a given time point with the data from the previous two weeks to create a 3-week moving average. We then ran retrospective forecasts as described in the main text using the smoothed data.

Retrospective Forecast Generation by Tropical Outbreak

To assess the role of sporadic outbreak timing on forecast accuracy in the tropics, we also ran tropical forecasts for each outbreak individually, similar to how forecasting was performed in temperate regions. Outbreaks for each tropical country were identified as described in the main text. We then added eight weeks to the beginning and end of each identified outbreak period, before performing forecasts as described in the main text for temperate regions, considering each outbreak as a "season." Specifically, fitting began eight weeks before the identified outbreak onset, and forecasts were generated starting 3 weeks before outbreak onset (corresponding to 5

weeks of training data, as in the temperate regions) through 4 weeks after the outbreak had ended. We note that, in temperate countries, model fitting started an average of 13 to 14 weeks prior to outbreak onset. However, because outbreaks in the tropics often happen in rapid succession, it was not possible to include this amount of lead time around the outbreak periods without also including substantial portions of other "outbreaks."

Sensitivity Analyses for Timing/Intensity Accuracy Cutoffs

Because our conclusions regarding forecast accuracy are dependent on the ranges of predicted timing and intensity values that we consider to be "accurate," we also assessed forecast accuracy using alternative accuracy definitions, one stricter and one more lenient. Specifically, we considered forecasts of peak timing to be accurate (a) only when the forecasted peak timing equaled the observed peak timing exactly, or (b) when the forecasted peak timing was within 2 weeks of the observed peak timing. For peak intensity, we considered forecasts accurate (a) when the forecasted value was within 12.5% of the observed value, or (b) when the forecasted value was within 50% of the observed value.

Alternative Forecast Accuracy Metrics

In addition to the peak timing, peak intensity, and onset timing accuracy metrics defined in the main text, we also assessed forecast accuracy over the duration of the forecast using correlation coefficients and the symmetric mean absolute percentage error (sMAPE). These metrics are calculated by comparing forecast influenza incidence from the time of forecast start until 10 weeks post peak with the observed influenza syndromic+ data over the same time period. This time period was chosen because, beyond 8 weeks post observed peak, syndromic+

case counts tend to be low, zero, or missing, precluding meaningful error measurements. We additionally removed any forecasts with fewer than four non-NA data points. sMAPE is defined as:

$$sMAPE = \frac{100\%}{T} \sum_{t=1}^{T} \frac{|F_t - O_t|}{(|O_t| + |F_t|)/\,2}$$

where T is the number of weeks forecasted, O_t is the observed syndromic+ value at time t, and F_t is the forecasted influenza incidence at time t (Tofallis, 2015). We chose to use sMAPE rather than the more commonly used root mean square error (RMSE) because, unlike RMSE, sMAPE controls for the difference in the magnitude of the observed data both at different points in an outbreak, as well as between different countries in the dataset. Also, unlike MAPE, sMAPE is not highly biased toward forecasts that undershoot observed values (Tofallis, 2015).

To test whether significant differences in forecast accuracy exist between temperate and tropical regions, we performed Kruskal-Wallis rank sum tests at predicted lead weeks -6 through 4. Because 5 individual runs were performed for each country and season, we randomly chose a single run for each country and outbreak combination 20 separate times, similar to the process described in the main text for comparing forecast accuracy. If p-values were below 0.0045 (0.05 / 11; p = 0.05 with Bonferroni correction for the 11 distinct lead weeks) for at least 50% of randomly selected run combinations, we considered there to be a significant difference in value for that lead week.

Method of Analogues

We further compared our mechanistic forecasting results with results obtained using the method of analogues (Viboud et al., 2003). Explicit methodological detail can be found in

Viboud et al. (2003). Briefly, the method involves searching through the entire time series of each country for a given number of vectors, or "nearest neighbors," that most closely match the data at the time at which a forecast is desired. We performed the method for each country individually at each time point using two nearest neighbors of length four. These nearest neighbors were drawn from previous seasons in the temperate regions, and from any previous data in the tropics; in other words, neighbors from the current season itself were not permitted when forecasting in the temperate regions. Additionally, because missing data were common in our dataset, we limited forecasting with the method of analogues to forecast start weeks where at least two of the preceding three weeks had data. Finally, as this method relies on patterns observed in past data, we do not begin forecasting until two full seasons (temperate) or outbreaks (tropics) have occurred. We note that this precludes forecasting in several tropical countries (Bangladesh, Bolivia, Brazil, Honduras, Indonesia, Kenya, and Madagascar). In order to fairly compare the method of analogues with our methods as described in the main text, we remove forecasts from our main results accordingly for this analysis only.

Data Quality Metrics

In general, we expect that forecast accuracy will be higher when data of better quality are used for model fitting. To test whether this was the case in this study, we calculated three measures of data quality:

1) The proportion of weeks within seasons (temperate countries; weeks 40 to 19 in the northern hemisphere and weeks 14 to 45 in the southern hemisphere) or outbreaks (tropical countries; outbreaks extended from the end of a previous outbreak to the next outbreak endpoint, so as to include both the outbreak itself and the most proximal

training data) for which no data were available were calculated overall for each country, as well as by season for temperate countries.

2) Data signal smoothness was calculated using lag-one autocorrelation.
3) The extent to which a country sampled for influenza was estimated by comparing the number of virologic samples taken each week within the influenza season (temperate) or throughout the year (tropics) to the country's total population size.

These measures were then compared to overall average peak timing and intensity accuracy by country (and by season, for temperate countries and measure 1) using Kendall's rank correlation.

Inferred Model States and Parameters

As described in the main text, model state variables (the number susceptible and infected) and parameters (R_{0max}, R_{0min}, R_0, D, L) are inferred throughout the model fitting process. To determine whether inferred values of *S0* (the initial number of susceptible individuals in a country) and of model parameters substantially differed between temperate and tropic countries, by hemisphere, or by data type, we first limited our analysis to the training period of the final forecasts run for each country and season, as these were the model fits incorporating the greatest number of data points. Then, we calculated R_0 for temperate countries according to equation 2 in the main text. Finally, R_e, or "R effective," defined as the number of cases caused by each infected individual after taking into account the susceptibility of the population, was calculated by multiplying R_0 by S / N at each time point. The value of *S0* for a country and season (or country and outbreak in the tropics) was considered to be the maximum inferred value of S over the course of the outbreak. R_0, R_e, D, and L were considered at the time of maximum R_e for each country and outbreak, as described in Yang, Lipsitch et al. (2015). Finally, values of *S0*, R_0, R_e,

D, and *L* were compared by region (temperate vs. tropics), hemisphere, and data type using the Kruskal-Wallis rank sum test, as described above under "Alternative Forecast Accuracy Methods." Here, results were considered significant if f p-values were below 0.05 for at least 50% of randomly selected run combinations.

Inferred Maximum and Minimum R_0 by Latitude

In each country, R_0 is allowed to vary between some maximum R_{0max} and some minimum R_{0min}, dependent on absolute humidity (see Eq. 2 in main text). R_{0max} and R_{0min} are fit separately for each country, and thus are permitted to vary. If the influence of humidity on influenza transmission acts similarly at all latitudes, we expect inferred values of R_{0max} and R_{0min} to also be similar at all latitudes.

To test this, we identified the inferred values of R_{0max} and R_{0min} for each country and season in both the northern and southern temperate regions at maximum R_e, as described in the previous section. We then compared values of these two parameters between hemispheres, as described above, as well as by latitude, using Kendall rank correlation. Again, results were considered significant if p-values were below 0.05 for at least 10 of 20 randomly selected combinations of model runs. For each country, we tested two different values for latitude: the latitude at the center of the country, and the latitude of the country's capital city. Absolute values were used so that countries in both the northern and southern hemispheres could be assessed together.

Pandemic Forecasts

We also generated retrospective forecasts for the 2009 influenza pandemic for the 34 countries (including 2 in the tropics) reporting data during this period. Because the time of pandemic emergence was not known in advance in real time, we did not begin forecasting until scaled observations exceeded 50% of the onset baseline value (250 in temperate regions and 150 in the tropics). At that point, an initial forecast was produced using 10 weeks of training data, and forecasting proceeded as described in the main text. For this reason, forecasts of pandemic onset were not possible. Note that if the time at which syndromic+ data exceeded baseline onset was before the fifth week of data, forecasting was begun after 2 weeks of training, to avoid generating forecasts with insufficiently trained models. This was done so that model states and parameters had some degree of training before forecasts were produced. Forecasts were then generated every week until scaled observations fell below 50% of the onset baseline for >=2 consecutive weeks (which we considered the pandemic "endpoint" for a country), or until less than 4 weeks remained before the beginning of the 2010-11 influenza season (in temperate countries). Thus, the exact period over which forecasts were generated varied by country. Otherwise, forecasts were run as described in the main text.

Countries for which both syndromic and virologic data were available for the 2009 pandemic included 32 northern hemisphere temperate countries, and 2 tropical countries (Honduras and Singapore). A complete list of these countries can be found in S2 Table.

Supplementary Results

Forecast Accuracy by Country

As observed in a previous forecasting study focusing on US cities (Shaman et al., 2013) and as mentioned in the main text, forecast accuracy varied greatly by individual country (S2 Figure).

Forecast Accuracy by Observed Lead Week

When assessed by observed lead week, retrospective forecasts for temperate regions reached 50% accurate from 5 weeks prior to the peak (for peak timing) and 1 week prior to the peak (for peak intensity). Peak timing forecasts exceed 75% at 2 weeks post peak, and peak intensity forecasts exceed 75% the week after the peak (S3 Figure A). These results are similar to those presented for predicted lead week in the main text.

Forecasts in the tropics exceed 50% accuracy at the observed peak for both peak timing and intensity (S3 Figure B). Forecasts surpass 75% accuracy at one week post peak for peak timing, but never reach 75% accuracy for peak intensity. Thus, results are similar to those in the main text before the peak, but demonstrate much higher accuracy post-peak.

Choice of OEV Denominator

S4 Figure compares peak timing and intensity forecast accuracy for temperate and tropical regions when c is set to 10, rather than 1 as in the main text, corresponding to a tenfold reduction of error in the syndromic+ observations. S5 Figure compares forecast calibration under the same circumstances. In temperate regions, setting c to 10 appears to have little impact on forecast accuracy. In the tropics, however, peak intensity accuracy appears substantially higher

when *c* is set to 10 rather than 1. However, for both temperate and tropical regions, and for peak intensity in particular, setting *c* to 1, as presented in the main text, appears to result in better forecast calibration, i.e. the prediction intervals for peak intensity are more aligned with the spread of observations. Given our goal of producing forecasts that are both accurate and well calibrated, using a *c* of 1 appears preferable to a *c* of 10.

Choice of Onset

In the main text, we set the scaled baseline value to 500 for temperate regions and 300 for the tropics. In S6 Figure, we present the overall accuracy of onset timing forecasts when onset is set to 300, 400, or 600 in the temperate regions, and 200, 400, or 500 in the tropics. Seasons where no onset occurred were removed, and forecasts predicting no onset were counted as inaccurate. Forecast accuracy is therefore presented by lead relative to observed onset week, as predicted lead onset week does not exist when no onset is predicted. In temperate regions, there are no substantial differences in onset timing forecast accuracy by choice of onset value. In the tropics, differences are somewhat more pronounced, but the overall structure of accuracy as a function of lead is similar; note that for some baseline values more spurious predictions of no onset were generated (the smaller dots sizes in S6 Figure).

Choice of Scaling Value

In our main analyses, we systematically selected the lowest scaling values that yielded overall attack rates between 15% and 50% of the model population for all seasons, where possible. Here, we test the sensitivity of our results to this decision by essentially flipping our

scaling selection rule (Eq. 4) and choosing the highest scaling values that yield the desired attack rates:

$$\gamma = \begin{Bmatrix} if\ \exists\ \gamma\ \in\ \mathbb{R}:\ \gamma_{15,i} < \gamma < \gamma_{50,i}\ \forall\ i: & min_{i=0}^{n}(\gamma_{50,i}) \\ else: & max_{i=0}^{n}(\gamma_{15,i}) \end{Bmatrix}$$

Results of these analyses are shown in S7 Figure. Overall, changing the selection rule has little impact on forecast accuracy.

Tropical Forecast Accuracy Using Smoothed Data

When forecasting in the tropics is performed using data smoothed with a 3-week moving average, forecast accuracy appears to improve slightly for peak intensity, but not for peak timing (S8 Figure). However, forecast accuracy remains much lower than in the temperate regions.

Retrospective Forecast Accuracy by Tropical Outbreak

When forecasts of tropical outbreaks are performed separately for each outbreak, essentially treating each outbreak as a "season," forecast accuracy before the predicted peak appears to increase slightly (S9 Figure). However, forecast accuracy remains low overall compared to temperate regions, suggesting that low forecast accuracy in the tropics is not primarily due to the irregularity of outbreaks, which prevent recurrent, seasonal model fitting and forecasting. Instead, the differences appear to be related to factors such as the high amount of noise in tropical observations.

Additional Forecast Calibration Results

In a properly calibrated forecast, we expect that errors in forecasted peak timing and intensity will display some distribution with a mean of 0. In contrast, a non-zero mean indicates that forecasts are biased. We assess whether the forecasts generated in the main text are biased according to this measure by plotting histograms of the difference between the observed and predicted peak timing and peak intensity over time (S10 Figure). In order to standardize errors over a wide range of observed peak intensity values by country, we plot the error in peak intensity forecasts divided by the observed peak intensity, rather than simply plotting the absolute error. Using this metric, we see that good calibration is achieved in the temperate regions, particularly directly prior to the peak. Calibration appears substantially worse in the tropics, where both peak timing and peak intensity are consistently underestimated. Thus, although forecasts of peak intensity in the tropics yield informative and well-constrained credible intervals (Figures 3 and 4 in the main text), they display substantial bias.

Forecast Accuracy Using Alternative Accuracy Cutoffs

As expected, calculated forecast accuracy decreased when stricter accuracy cutoffs were employed (S11 Figure A and B) and increased when less strict cutoffs were used (S11 Figure C and D), However, observed patterns in accuracy remained the same: forecast accuracy generally increased as predicted lead week increased, and accuracy in temperate regions was noticeably higher than in the tropics.

Forecast Accuracy Using Correlation Coefficients and sMAPE

Correlations between observed and forecasted incidence were significantly higher for temperate than tropical countries for all lead weeks except predicted lead week -6, where very few forecasts were available (S12 Figure A and B). Also notable were the wide confidence intervals around correlation coefficient estimates in the tropics, with 95% credible intervals ranging from -0.68 to 0.97 (as opposed to 0.46 to 0.995 in temperate countries). sMAPE values, on the other hand, were similar in temperate and tropical regions, with no statistically significant differences observed at any predicted lead week. While we believe that the targets used in the main text (peak timing and peak intensity) represent metrics of practical importance for responding to influenza outbreaks, it is nonetheless important to acknowledge that the impact of temperate vs. tropical region on forecast accuracy is dependent on how forecast accuracy is measured, and that tropical forecasts may perform better for other metrics not measured here.

Method of Analogues Forecast Accuracy

Results are primarily discussed in the main text. However, we note here that, because the method of analogues requires information on past outbreaks, early outbreaks could not be forecasted for any country, and several tropical countries had to be removed from consideration entirely. Before comparing to mechanistic forecasts (S13 Figure), we therefore removed any country or season that could not be forecasted using the method of analogues.

Forecast Accuracy by Data Quality

All three measures of data quality were found to differ significantly between temperate and tropical regions (Kruskal-Wallis one-way analysis of variance, $p < 0.01$ for all measures).

Therefore, the relationship between these measures and forecast accuracy was assessed separately for temperate and tropical countries.

Greater smoothness of data signal was significantly associated with higher peak intensity accuracy among both temperate (Kendal's tau = 0.262, $p < 0.05$) and tropical (Kendal's tau = 0.464, $p < 0.01$) countries, but not with peak timing accuracy. Neither proportion of data missing nor proportion of population sampled was significantly associated with forecast accuracy in either the temperate or tropical regions.

Models States and Parameters

Broadly, inferred states and parameters fall within realistic ranges (Carrat & Flahault, 2007; Mills et al., 2004; Truscott et al., 2009; White & Pagano, 2008), with values between about 50% and 90% of the population for *S0*, 1.0 and 5.3 for R_e, 1.4 and 3.1 for R_0, 2.3 to 8.4 days for *D*, and 3.8 to 7.7 years for *L*. Compared to temperate countries, countries in the tropics yielded significantly lower values of *S0*, R_0, R_e, and *L* (S14 Figure A). Within temperate regions, countries located in the northern hemisphere showed significantly higher values of both R_e and R_0 than southern hemisphere countries (S14 Figure B), and countries and seasons reporting ILI+ data displayed significantly lower R_0 than countries reporting ARI+ data (S14 Figure C). No significant differences were observed between data types in the tropics (S14 Figure D).

Inferred Maximum and Minimum R_0 by Latitude

Neither R_{0max} nor R_{0min} varied significantly by hemisphere, but R_{0max} was significantly and negatively associated with the absolute value of latitude for both definitions of latitude used (S15 Figure A and B). In other words, as the distance from the equator increased, the maximum

possible *R0* tended to decrease, suggesting a weaker impact of absolute humidity at higher latitudes. That said, the relationships were weak, with Kendall's tau ranging from -0.05 to -0.10 when capital cities' latitudes were used, and from -0.10 to -0.15 when centroids were used. Such a nominal result is more likely due to the simplicity of our model or the large geographic scale at which our model is implemented, than to a true biological process.

Posterior and Forecast Visualizations

S16 Figure shows posterior model fits for five countries: Norway, Poland, Italy, Mexico, and Ecuador. These countries were chosen because they inhabit a range of latitudes and longitudes, and exhibited similar peak weeks (week 8 of 2016 for the four temperate countries, and week 17 of the same year for Ecuador). Furthermore, all four temperate countries reported ILI data. For the temperate countries, because the model was fit separately for each season, only the 2015-16 season was plotted. For Ecuador, which is located in the tropics model fitting is shown throughout the entirety of the time series leading up to the peak of interest. The mean posterior was plotted for all five model runs. As can be observed, the model was capable of closely fitting the data for a range of countries with varying locations and climates.

S17 Figure and S18 Figure show forecast trajectories over several lead weeks for the same five countries (the four temperate countries in S17 Figure, and Ecuador in S18 Figure). Both peak timing and intensity were predicted within 1 week and within 25% of observed values, respectively, by 2 weeks prior to the observed peak in Norway, Poland, and Italy, but were not both predicted accurately in Mexico until 2 weeks post-peak. In Ecuador, as was common for countries in the tropics, neither peak timing nor peak intensity were accurately predicted until the peak occurred. Additionally, the model was unable to detect the epidemic signal, with

trajectories consistently predicting decreasing rather than increasing incidence, even before the peak.

Pandemic Forecast Accuracy

In temperate regions, pandemic forecasts appear to perform slightly worse that seasonal forecasts prior to the predicted peak for both peak timing and intensity, with peak timing first exceeding 50% accuracy 2 weeks before the predicted peak, and peak intensity not exceeding 50% accuracy until the predicted peak (S19 Figure A). Given that the pandemic often did not display the clear signal and single peak typical of regular seasonal outbreaks, this finding is not surprising. Also, note that pandemic forecasting often had to be begun abruptly when out-of-season increases in influenza activity were observed.

In the tropics, forecasts of pandemic peak timing were more accurate than similar forecasts for "seasonal" influenza outbreaks several weeks before the predicted peak, and post-peak estimates of pandemic peak intensity also appeared better than analogous estimates for epidemic influenza (S19 Figure B). However, it is important to note that, because pandemic data were only available for 2 tropical countries, forecast counts are very low, reducing the certainty of these results.

References

Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. *Vaccine, 25*(39–40), 6852–6862. https://doi.org/10.1016/j.vaccine.2007.07.027

Mills, C. E., Robins, J. M., & Lipsitch, M. (2004). Transmissibility of 1918 pandemic influenza. *Nature*, *432*(7019), 904–906. https://doi.org/10.1038/nature03063

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications, 4*. https://doi.org/10.1038/ncomms3837

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, *66*(8), 1352–1362. https://doi.org/10.1057/jors.2014.103

Truscott, J., Fraser, C., Hinsley, W., Cauchemez, S., Donnelly, C., Ghani, A., Ferguson, N., & Meeyai, A. (2009). Quantifying the transmissibility of human influenza and its seasonal variation in temperate regions. *PLoS Currents*, *1*, RRN1125. https://doi.org/10.1371/currents.RRN1125

Viboud, C., Boëlle, P.-Y., Carrat, F., Valleron, A.-J., & Flahault, A. (2003). Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, *158*(10), 996–1006.

White, L. F., & Pagano, M. (2008). Transmissibility of the influenza virus in the 1918 pandemic. *PloS One*, *3*(1), e1498. https://doi.org/10.1371/journal.pone.0001498

WHO. (2011). *WHO Regional Office for Europe guidance for sentinel influenza surveillance in humans*. http://www.euro.who.int/__data/assets/pdf_file/0020/90443/E92738.pdf

Yang, W., Cowling, B. J., Lau, E. H. Y., & Shaman, J. (2015). Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology*, *11*(7), e1004383. https://doi.org/10.1371/journal.pcbi.1004383

Yang, W., Lipsitch, M., & Shaman, J. (2015). Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, *112*(9), 2723–2728. https://doi.org/10.1073/pnas.1415012112

Supplementary Tables

S1 Table. Countries used for retrospective forecasting, by region, data type, and scaling.

Country	Hemisphere	Region	Data Type	Scaling
Austria	North	Southwest Europe	ILI	14
Belgium	North	Southwest Europe	ILI	8.5
Croatia	North	Southwest Europe	ILI	1.4
France	North	Southwest Europe	ARI/ILI	1.3/0.03
Germany	North	Southwest Europe	ARI	0.25
Greece	North	Southwest Europe	ILI	4
Italy	North	Southwest Europe	ILI	1
Luxembourg	North	Southwest Europe	ARI	32
Netherlands	North	Southwest Europe	ILI	31
Portugal	North	Southwest Europe	ILI	245
Serbia	North	Southwest Europe	ILI	0.5
Slovenia	North	Southwest Europe	ARI	5.8
Spain	North	Southwest Europe	ILI	2
Belarus	North	Eastern Europe	ARI	0.2
Bulgaria	North	Eastern Europe	ARI	1.25
Czechia	North	Eastern Europe	ILI	0.8
Georgia	North	Eastern Europe	ILI	15
Hungary	North	Eastern Europe	ILI	1
Israel	North	Eastern Europe	ILI	3.5
Kazakhstan	North	Eastern Europe	ARI	0.35
Kyrgyzstan	North	Eastern Europe	ARI	0.63
Poland	North	Eastern Europe	ILI	1.3
Republic of Moldova	North	Eastern Europe	ARI	2.25
Romania	North	Eastern Europe	ILI	39
Russian Federation	North	Eastern Europe	ARI	0.02
Slovakia	North	Eastern Europe	ILI	0.57
Turkey	North	Eastern Europe	ILI	3.5
Ukraine	North	Eastern Europe	ARI	0.03
Uzbekistan	North	Eastern Europe	ARI	68
Denmark	North	Northern Europe	ILI	18
Estonia	North	Northern Europe	ARI	1.5
Finland	North	Northern Europe	ILI	14
Iceland	North	Northern Europe	ILI	57
Ireland	North	Northern Europe	ILI	45
Latvia	North	Northern Europe	ARI	4.4
Lithuania	North	Northern Europe	ILI	1.3
Norway	North	Northern Europe	ILI	1.65

Sweden	North	Northern Europe	ILI	132
United Kingdom	North	Northern Europe	ARI	1.8
Canada	North	Northern Hemisphere (non-Europe)	ILI Rate	182
Mexico	North	Northern Hemisphere (non-Europe)	ARI	0.004
Morocco	North	Northern Hemisphere (non-Europe)	ILI	0.78
United States of America	North	Northern Hemisphere (non-Europe)	ILI	0.27
Australia	South	Southern Hemisphere	ILI	32
Chile	South	Southern Hemisphere	ILI	31
New Zealand	South	Southern Hemisphere	ILI	16
Brazil	Tropics	Latin America	SARI	6
Bolivia	Tropics	Latin America	ARI	0.03
Colombia	Tropics	Latin America	ARI	0.075
Cuba	Tropics	Latin America	ILI	127
Ecuador	Tropics	Latin America	SARI	48
Honduras	Tropics	Latin America	Pneumonia	2
Paraguay	Tropics	Latin America	ILI	3.5
Peru	Tropics	Latin America	SARI	374
Kenya	Tropics	Africa/Middle East	ILI	10
Madagascar	Tropics	Africa/Middle East	ILI	6.6
Oman	Tropics	Africa/Middle East	SARI	6.4
Pakistan	Tropics	Africa/Middle East	ILI	93
Bangladesh	Tropics	South/Southeast Asia	SARI	48
Bhutan	Tropics	South/Southeast Asia	ILI	12
Cambodia	Tropics	South/Southeast Asia	SARI	90
Indonesia	Tropics	South/Southeast Asia	ILI	26
Singapore	Tropics	South/Southeast Asia	ILI	4.7
Thailand	Tropics	South/Southeast Asia	ILI	0.04

S2 Table. Countries and seasons used for retrospective forecasting. Countries used in retrospective forecasting of the 2009 pandemic are marked under "09pdm."

Country	09pdm	10-11	11-12	12-13	13-14	14-15	15-16	16-17
Austria		x	x	x	x	x	x	x
Belgium	x	x	x	x	x	x	x	x
Croatia	x	x	x	x	x	x	x	x
France				x	x	x	x	x
Germany	x	x	x	x	x	x	x	x
Greece	x	x	x	x	x	x		x
Italy	x	x	x	x	x	x	x	x
Luxembourg	x	x	x	x	x	x	x	x
Netherlands	x	x	x	x	x	x	x	x
Portugal		x	x	x	x	x	x	x
Serbia	x	x		x	x	x	x	
Slovenia	x	x	x	x	x	x	x	x
Spain	x	x	x	x	x	x	x	x
Belarus		x	x	x	x	x	x	x
Bulgaria	x	x	x	x	x	x	x	x
Czechia	x	x	x	x		x	x	x
Georgia	x	x	x	x	x	x	x	x
Hungary	x	x	x	x	x	x	x	x
Israel	x	x	x	x	x	x	x	x
Kazakhstan		x	x	x	x	x	x	x
Kyrgyzstan	x	x	x	x		x	x	x
Poland	x	x		x	x	x	x	x
Republic of Moldova	x	x	x	x	x	x	x	x
Romania	x	x	x	x	x	x	x	x
Russian Federation	x	x	x	x	x	x	x	x
Slovakia	x	x	x	x	x	x	x	x
Turkey	x	x	x	x	x	x	x	x
Ukraine	x	x	x	x	x	x	x	x
Uzbekistan				x	x	x	x	x
Denmark	x	x	x	x	x	x	x	x
Estonia	x	x	x	x	x	x	x	x
Finland				x	x	x	x	x
Iceland		x	x	x	x	x	x	x
Ireland		x	x	x	x	x	x	x
Latvia	x	x	x	x	x	x	x	x
Lithuania		x		x		x	x	x
Norway	x	x	x	x		x	x	x
Sweden	x	x	x	x	x			
United Kingdom	x	x	x	x	x			
Canada	x	x	x	x	x		x	
Mexico			x	x	x	x	x	x
Morocco	x	x	x	x		x	x	x
United States of America	x	x	x	x	x	x	x	x
Australia			x	x	x	x	x	x
Chile							x	x
New Zealand			x	x	x	x	x	

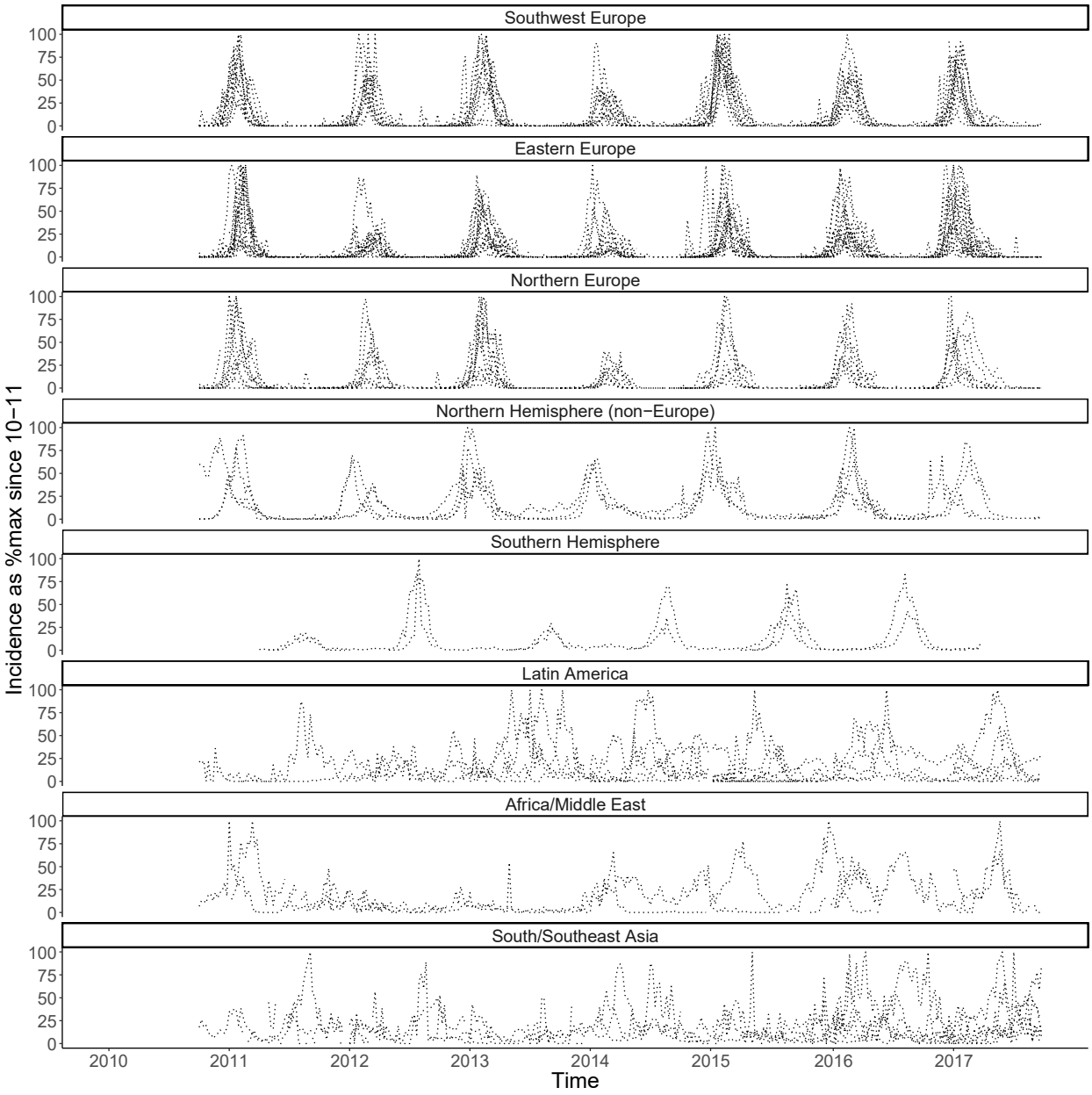
S3 Table. Peak timing and intensity accuracy overall, before the predicted peak, and at or after the predicted peak in temperate regions by hemisphere, season, region, data type, and scaling. Cells shaded in green indicate improved forecast accuracy over the reference level, while cells shaded in red indicate reduced accuracy.

Variable		Peak Timing aOR (95% CI)			Peak Intensity aOR (95% CI)		
		Overall	Before Peak	After Peak	Overall	Before Peak	After Peak
Hemisphere	Southern	2.125 (0.421, 10.684)	1.428 (0.562, 3.599)	1.134 (0.382, 3.384)	0.693 (0.269, 1.765)	0.940 (0.432, 2.052)	0.436 (0.108, 1.725)
Season	2016-17	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	2015-16	1.252 (0.603, 2.624)	1.105 (0.552, 2.211)	0.670 (0.269, 1.693)	0.795 (0.450, 1.401)	0.764 (0.437, 1.335)	0.761 (0.223, 2.570)
	2014-15	0.997 (0.481, 2.079)	1.004 (0.470, 2.127)	1.274 (0.434, 3.736)	1.071 (0.630, 1.836)	0.958 (0.543, 1.678)	1.289 (0.327, 4.810)
	2013-14	0.963 (0.441, 2.134)	0.752 (0.347, 1.636)	0.781 (0.304, 2.014)	1.915 (0.991, 3.750)	1.841 (0.969, 3.493)	Inf (0.292, Inf)
	2012-13	1.047 (0.499, 2.200)	1.177 (0.570, 2.437)	0.985 (0.365, 2.668)	1.115 (0.638, 1.950)	1.145 (0.640, 2.030)	0.701 (0.214, 2.330)
	2011-12	0.933 (0.433, 2.013)	0.911 (0.419, 1.980)	1.141 (0.391, 3.332)	1.433 (0.869, 2.341)	1.583 (0.925, 2.678)	1.431 (0.324, 6.128)
	2010-11	1.540 (0.726, 3.306)	1.333 (0.664, 2.665)	0.783 (0.298, 2.062)	1.033 (0.619, 1.726)	0.911 (0.526, 1.577)	1.610 (0.383, 7.018)
Region	SW Europe	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	E Europe	1.216 (0.737, 1.999)	1.045 (0.645, 1.690)	2.068 (1.095, 3.889)	0.744 (0.519, 1.067)	0.770 (0.527, 1.126)	0.962 (0.431, 2.126)
	N Europe	1.284 (0.691, 2.401)	1.286 (0.714, 2.324)	1.515 (0.718, 3.199)	1.114 (0.775, 1.605)	1.039 (0.674, 1.609)	Inf (1.034, Inf)
	N Hem (non-Europe)	0.807 (0.380, 1.736)	0.844 (0.403, 1.750)	2.036 (0.726, 5.745)	1.713 (0.835, 3.453)	1.686 (0.839, 3.356)	2.700 (0.458, Inf)
	S Hem	2.331 (0.471, 11.672)	1.475 (0.556, 3.955)	1.716 (0.548, 5.377)	0.656 (0.247, 1.743)	0.905 (0.403, 2.055)	0.548 (0.122, 2.386)
Data Type	ARI+	0.639 (0.421, 0.968)	0.645 (0.428, 0.973)	0.567 (0.334, 0.965)	0.832 (0.606, 1.132)	0.776 (0.562, 1.084)	0.741 (0.325, 1.721)
Scaling	(2, 10]	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	(0, 0.5]	0.771 (0.399, 1.488)	0.920 (0.473, 1.783)	0.420 (0.181, 0.982)	0.633 (0.397, 1.007)	0.846 (0.498, 1.421)	0.170 (0.051, 0.568)
	(0.5, 1]	1.361 (0.636, 2.941)	1.452 (0.701, 3.008)	1.758 (0.618, 5.075)	0.862 (0.524, 1.412)	1.044 (0.599, 1.818)	0.374 (0.100, 1.511)
	(1, 2]	1.360 (0.705, 2.612)	1.379 (0.704, 2.686)	1.815 (0.641, 5.115)	0.957 (0.648, 1.409)	1.058 (0.631, 1.765)	2.387 (0.412, 13.290)
	(10, 20]	1.869 (0.729, 4.785)	1.441 (0.667, 3.086)	0.799 (0.286, 2.244)	0.681 (0.401, 1.153)	1.115 (0.626, 1.990)	0.145 (0.038, 0.571)
	(20, 100]	0.863 (0.449, 1.650)	0.932 (0.483, 1.801)	0.667 (0.266, 1.674)	0.609 (0.374, 1.004)	0.873 (0.519, 1.463)	0.229 (0.066, 0.801)
	(100, 500]	1.494 (0.516, 4.120)	1.227 (0.520, 2.857)	0.956 (0.274, 3.336)	0.479 (0.250, 0.909)	0.604 (0.276, 1.345)	0.147 (0.033, 0.646)

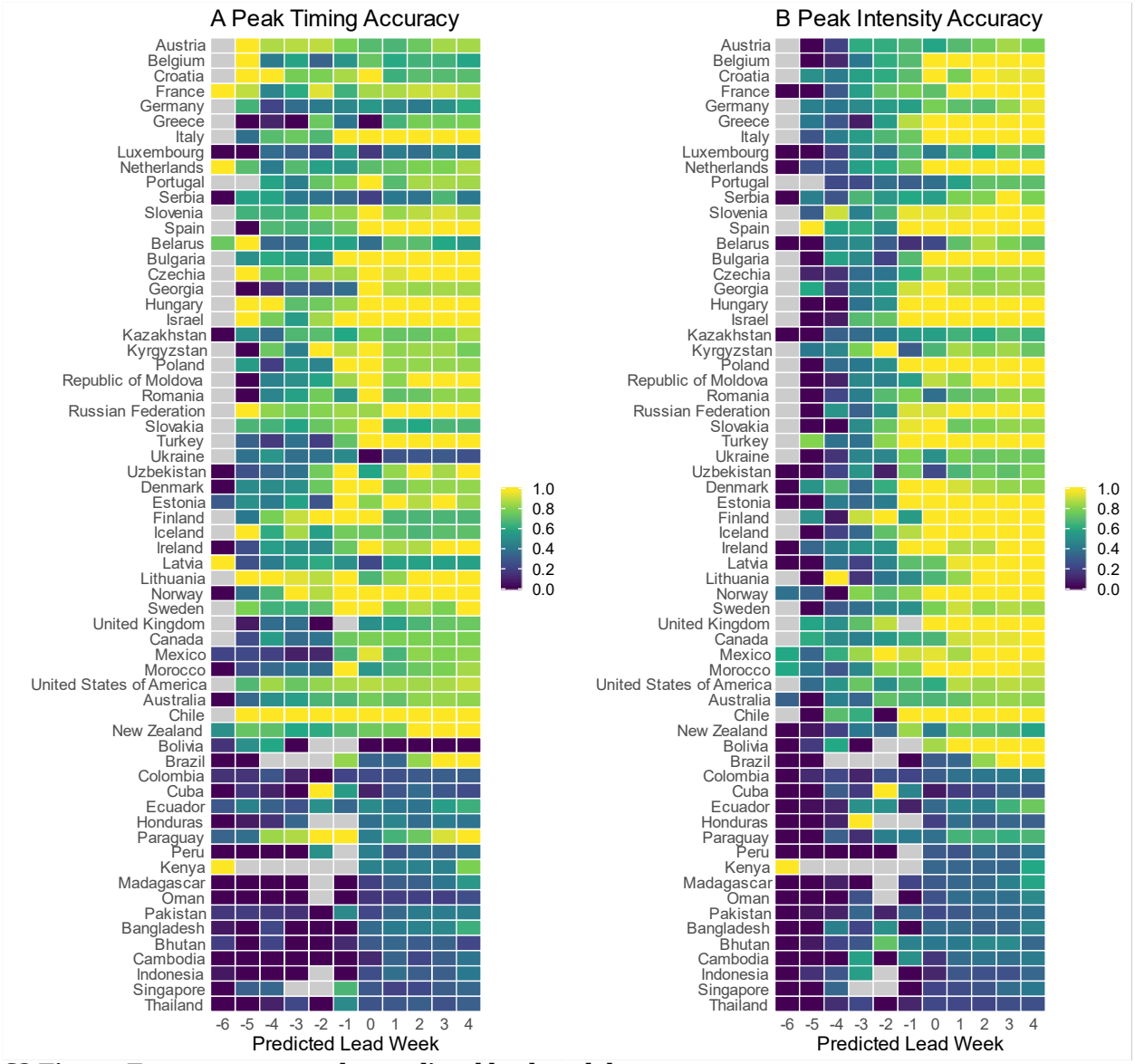
S4 Table. Peak timing and intensity accuracy at or after the predicted peak in the tropics by region, data type, and scaling. Cells shaded in green indicate improved forecast accuracy over the reference level, while cells shaded in red indicate reduced accuracy.

Variable		Peak Timing aOR (95% CI)	Peak Intensity aOR (95% CI)
Region	Latin America	1.00 (ref)	1.00 (ref)
	N Africa/Middle East	0.730 (0.420, 1.279)	0.751 (0.412, 1.374)
	SE Asia	0.722 (0.491, 1.057)	0.799 (0.425, 1.516)
Data Type	ILI+	1.00 (ref)	1.00 (ref)
	ARI+	0.822 (0.505, 1.341)	1.772 (0.726, 4.335)
	SARI+	1.158 (0.702, 1.904)	1.182 (0.663, 2.109)
	Pneumonia+	1.042 (0.756, 1.420)	0.524 (0.207, 1.613)
Scaling	(2, 10]	1.00 (ref)	1.00 (ref)
	(0, 2]	0.902 (0.528, 1.539)	0.899 (0.411, 1.942)
	(10, 20]	1.239 (0.707, 2.185)	1.321 (0.704, 2.471)
	(50, 300]	1.012 (0.556, 1.862)	0.700 (0.328, 1.473)

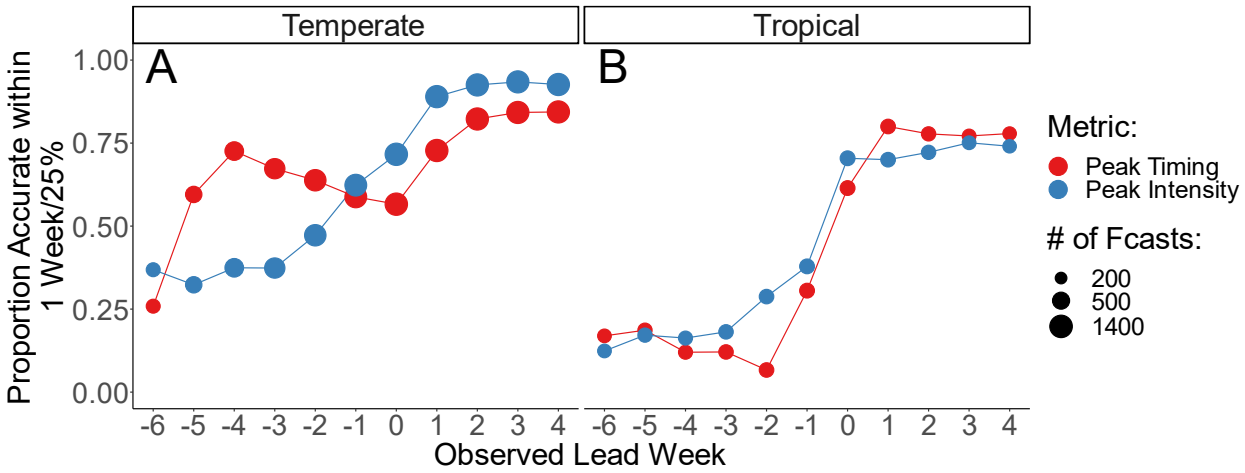
Supplementary Figures



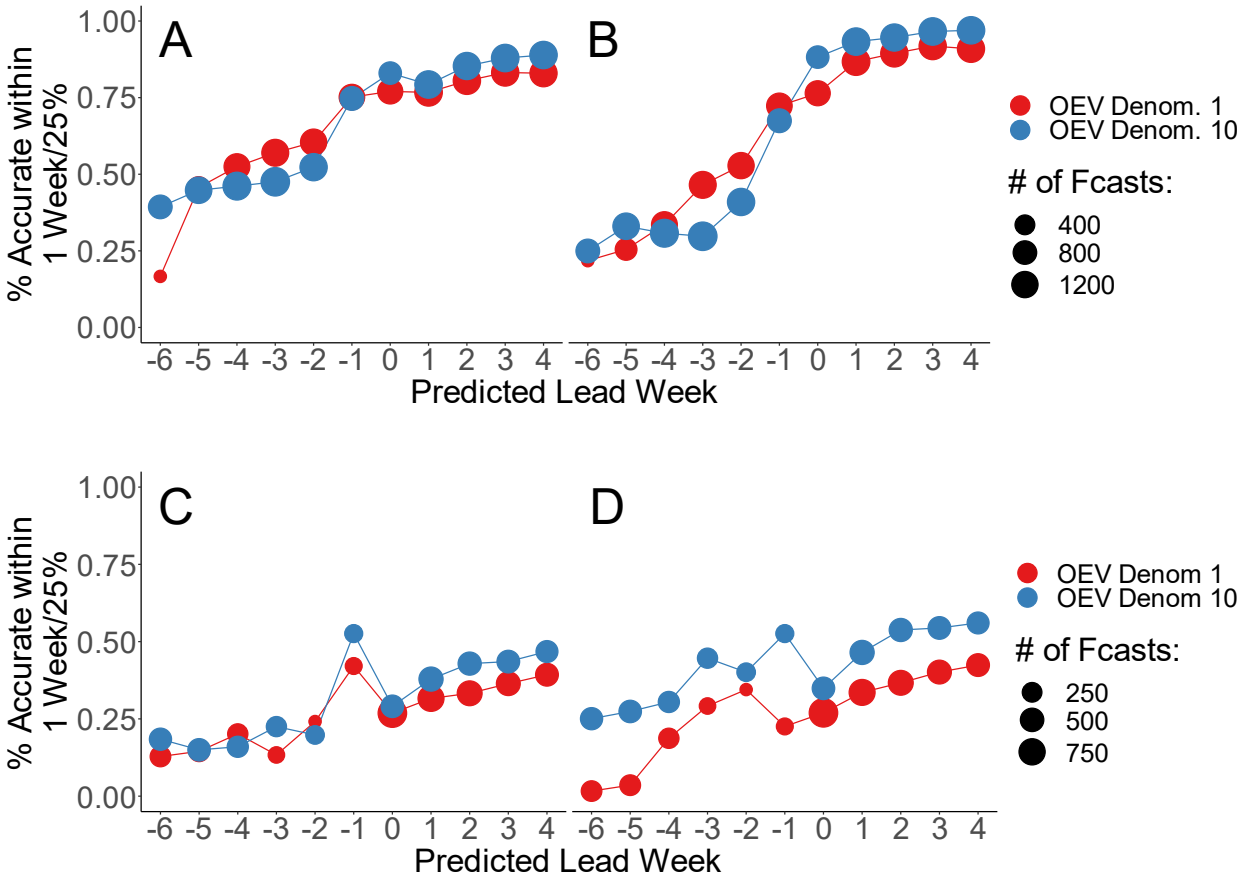
S1 Figure. Syndromic+ data by region.
All data are divided by the maximum observed incidence in a given country since the 2010-11 season.



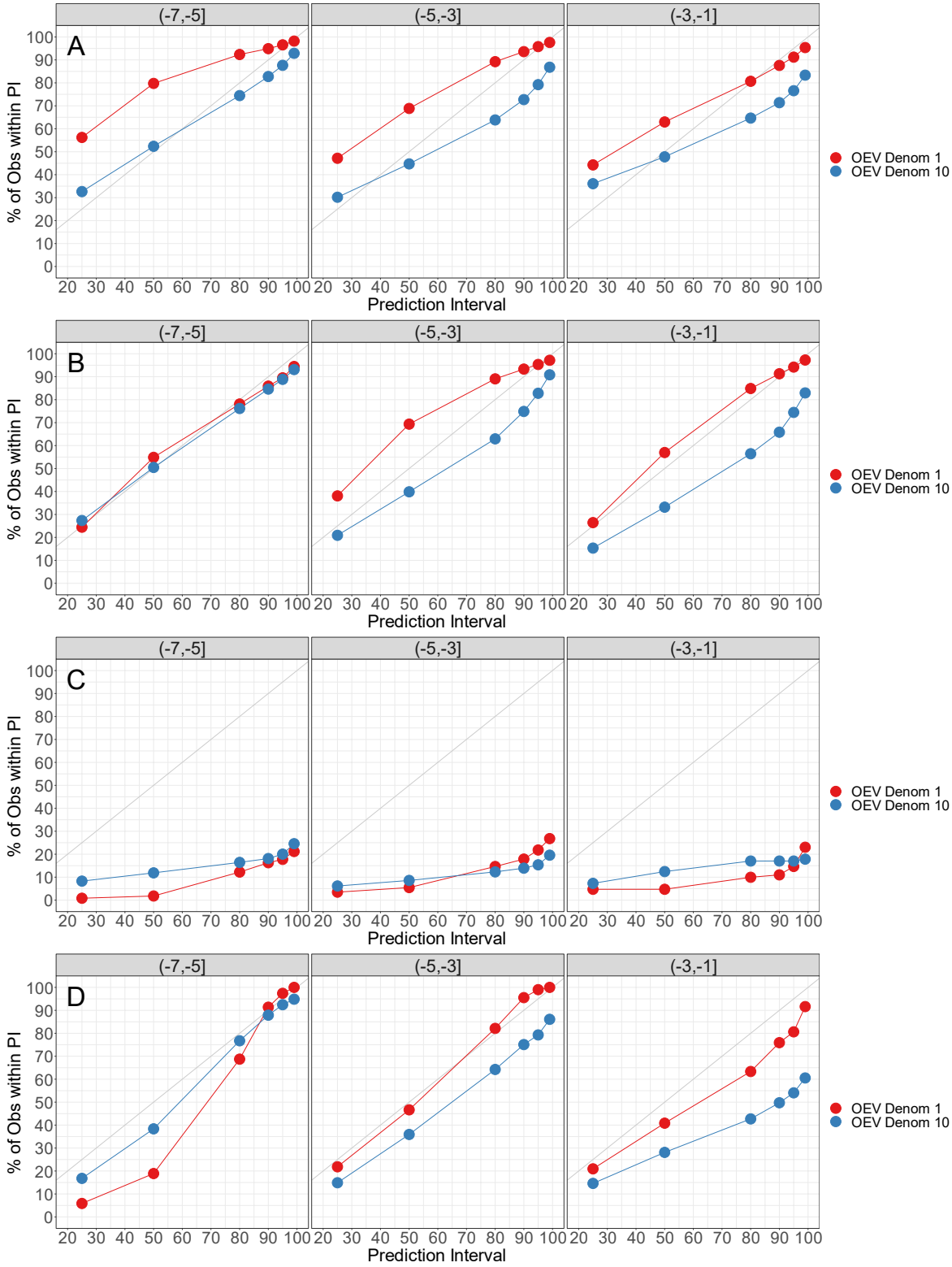
S2 Figure. Forecast accuracy by predicted lead week by country.
(A) Peak timing accuracy. (B) Peak intensity accuracy. NA values are represented by gray boxes.



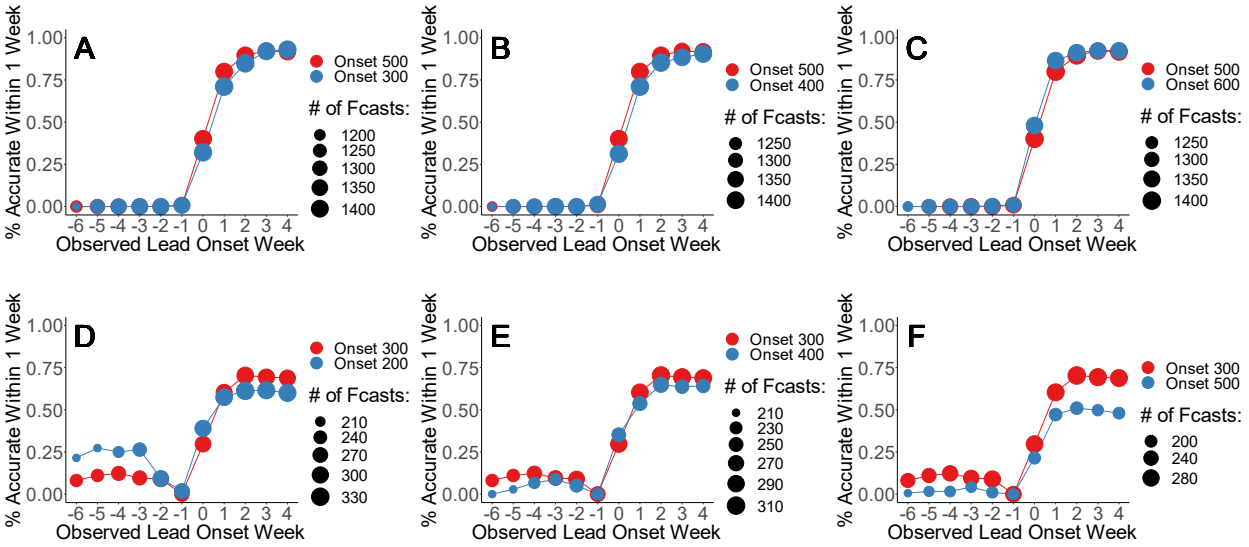
S3 Figure. Peak timing and intensity forecast accuracy by observed lead week.
(A) Forecast accuracy in temperate regions. (B) Forecast accuracy in tropical regions. Peak timing accuracy is shown in red, and peak intensity in blue.



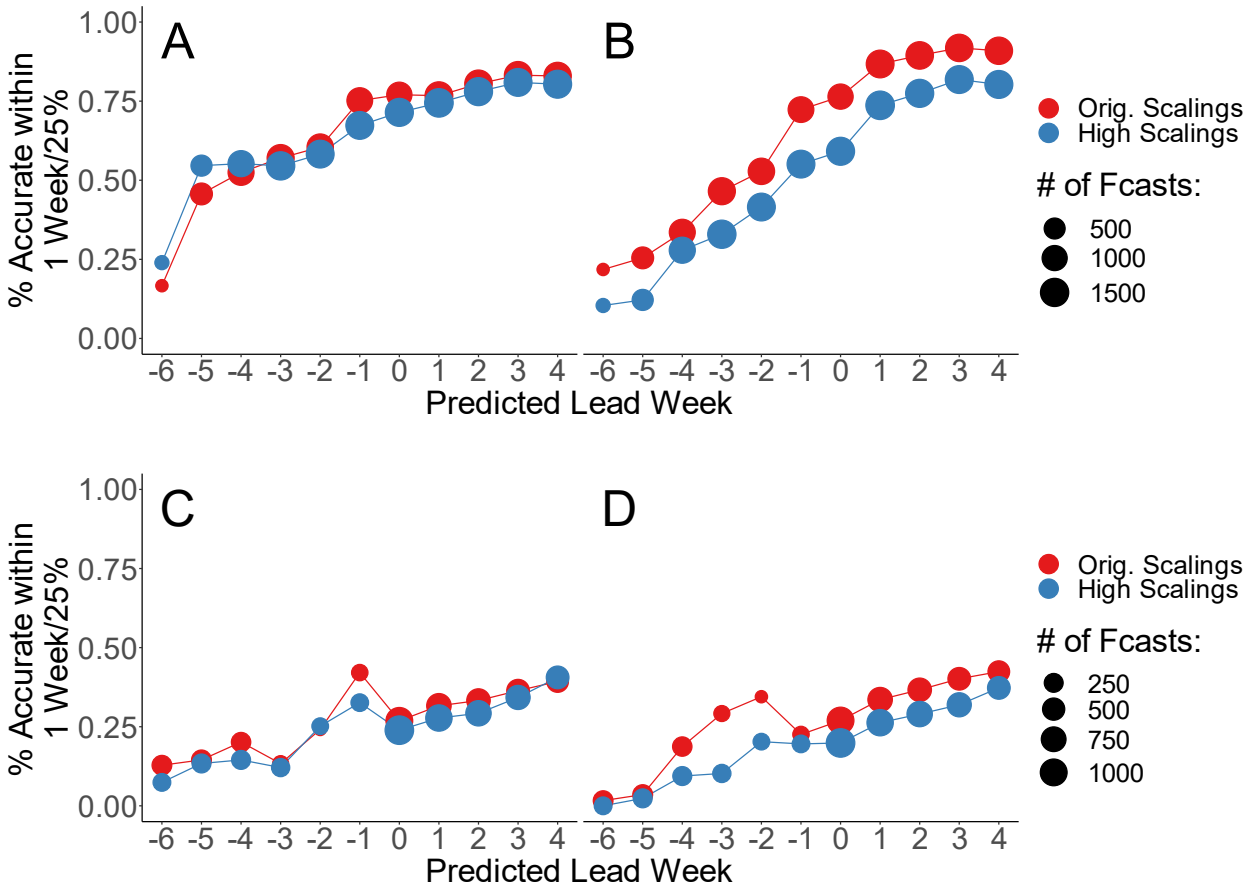
S4 Figure. Forecast accuracy by OEV denominator choice.
Peak timing (A and C) and intensity (B and D) accuracy for different OEV denominators in temperate (A and B) and tropical (C and D) regions.



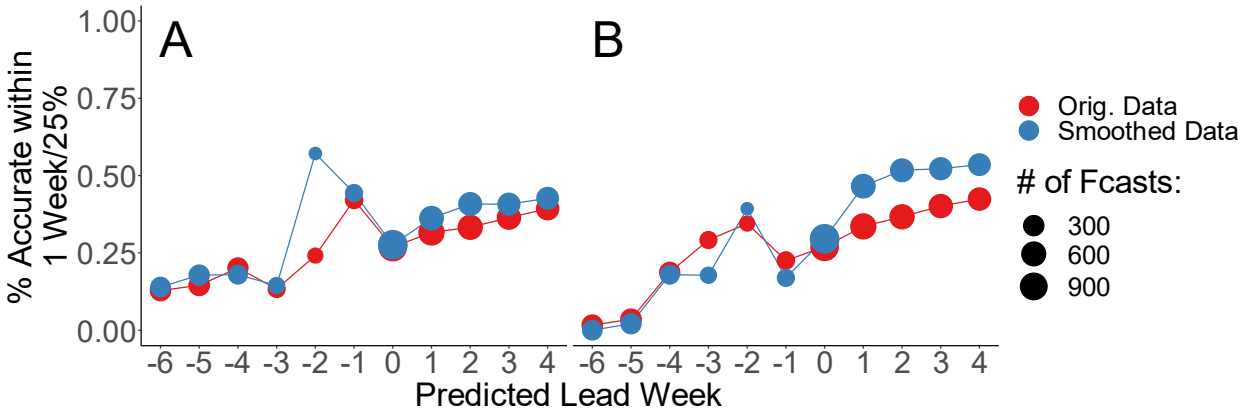
S5 Figure. Forecast calibration by OEV denominator choice.
Peak timing (A and C) and intensity (B and D) calibration by OEV denominator in temperate (A and B) and tropical (C and D) regions.



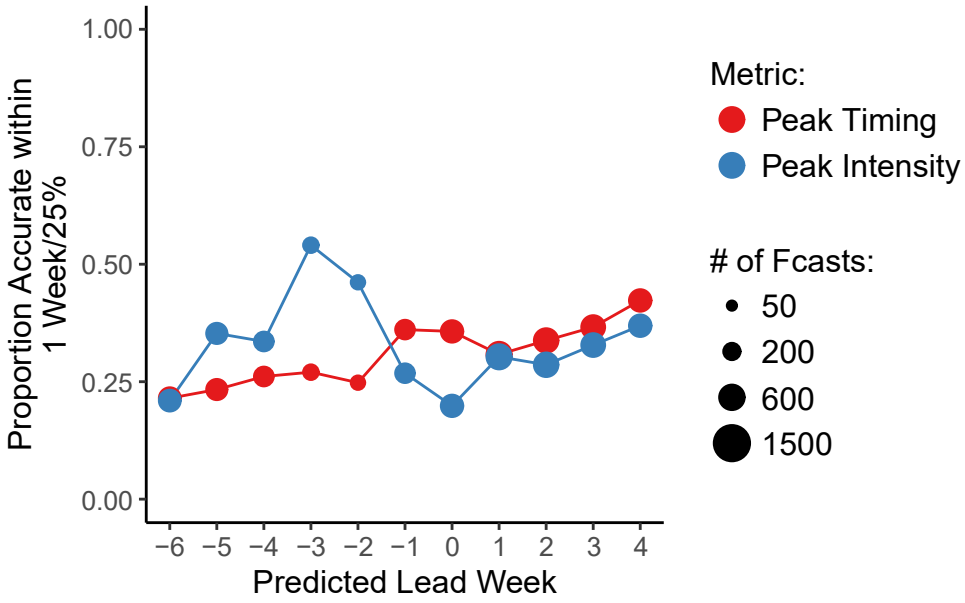
S6 Figure. Forecast accuracy by choice of onset value.
Onset timing accuracy by choice of onset value in temperate (A-C) and tropical (D-F) regions.



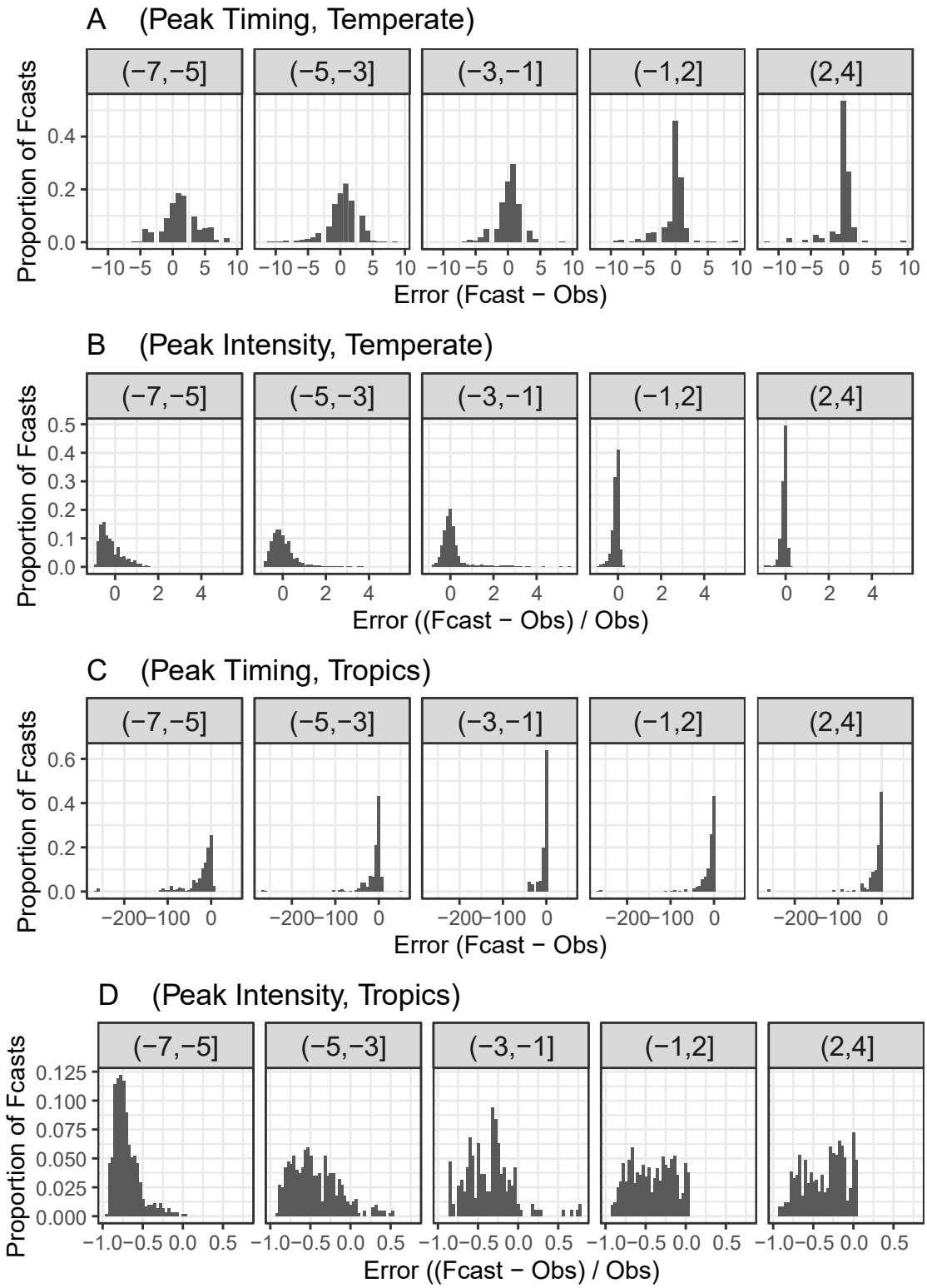
S7 Figure. Forecast accuracy by choice of scaling rule.
Peak timing (A and C) and intensity (B and D) accuracy by choice of scaling rule in temperate (A and B) and tropical (C and D) regions.



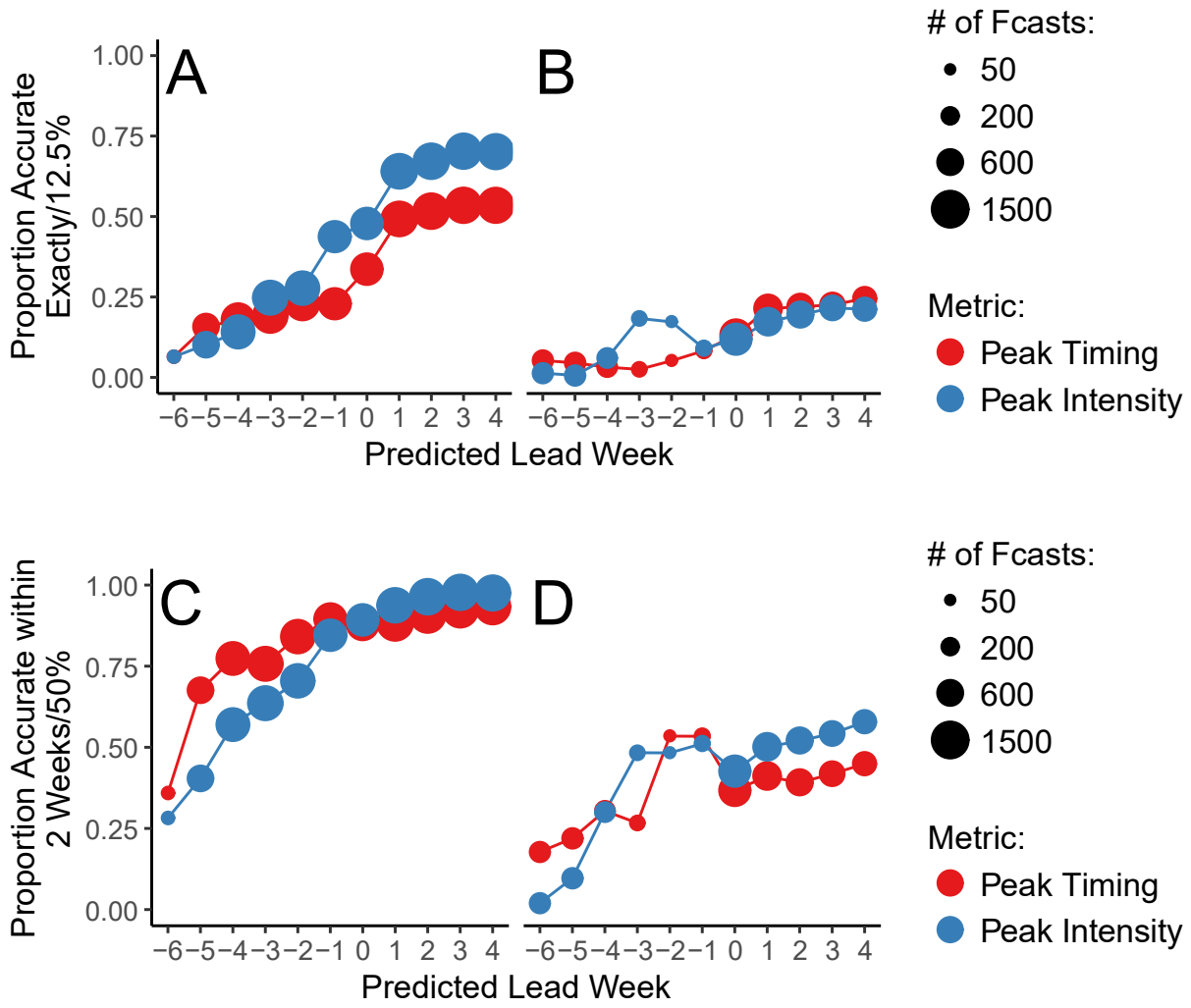
S8 Figure. Forecast accuracy for the tropics using smoothed and unsmoothed data.
(A) Peak timing accuracy. (B) Peak intensity accuracy.



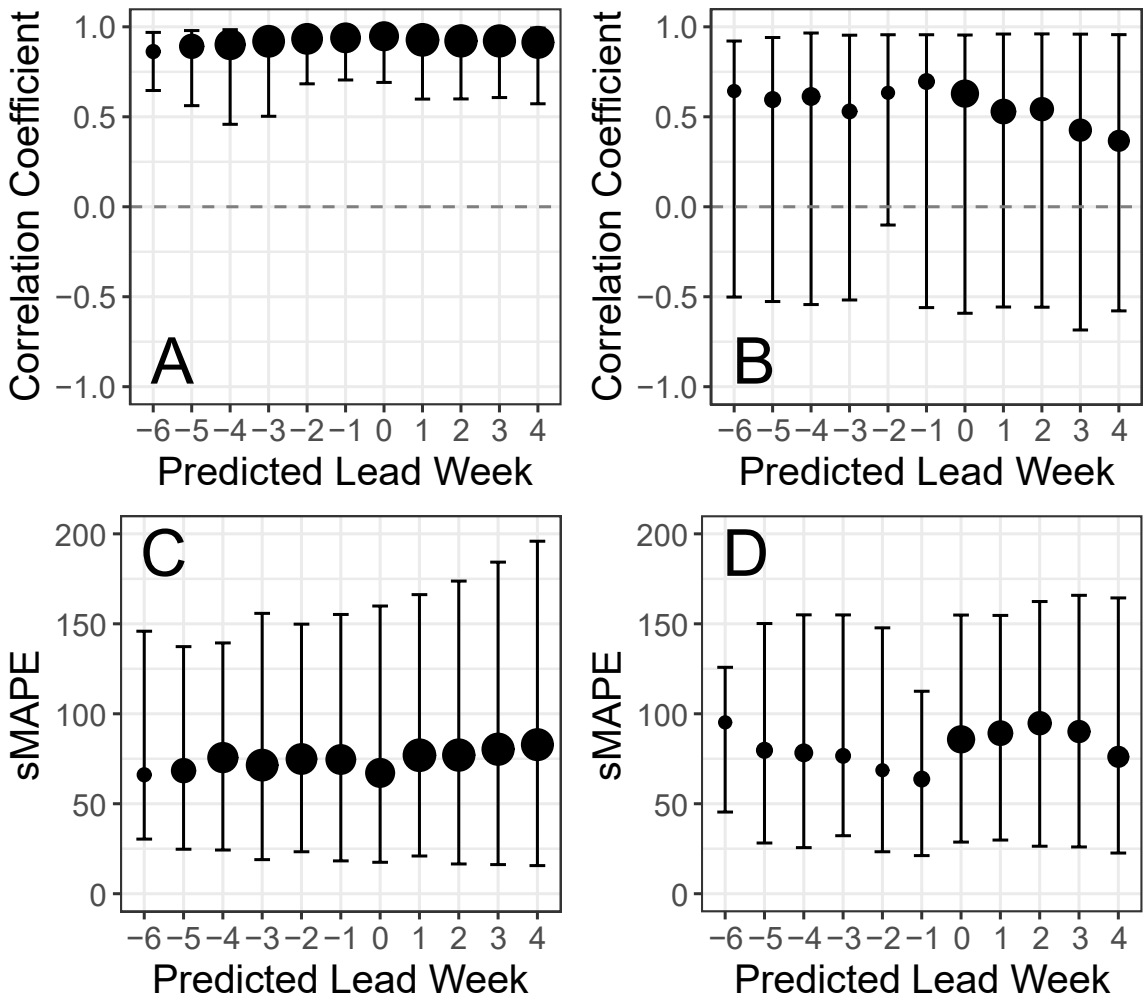
S9 Figure. Forecast accuracy for individual tropical outbreaks.



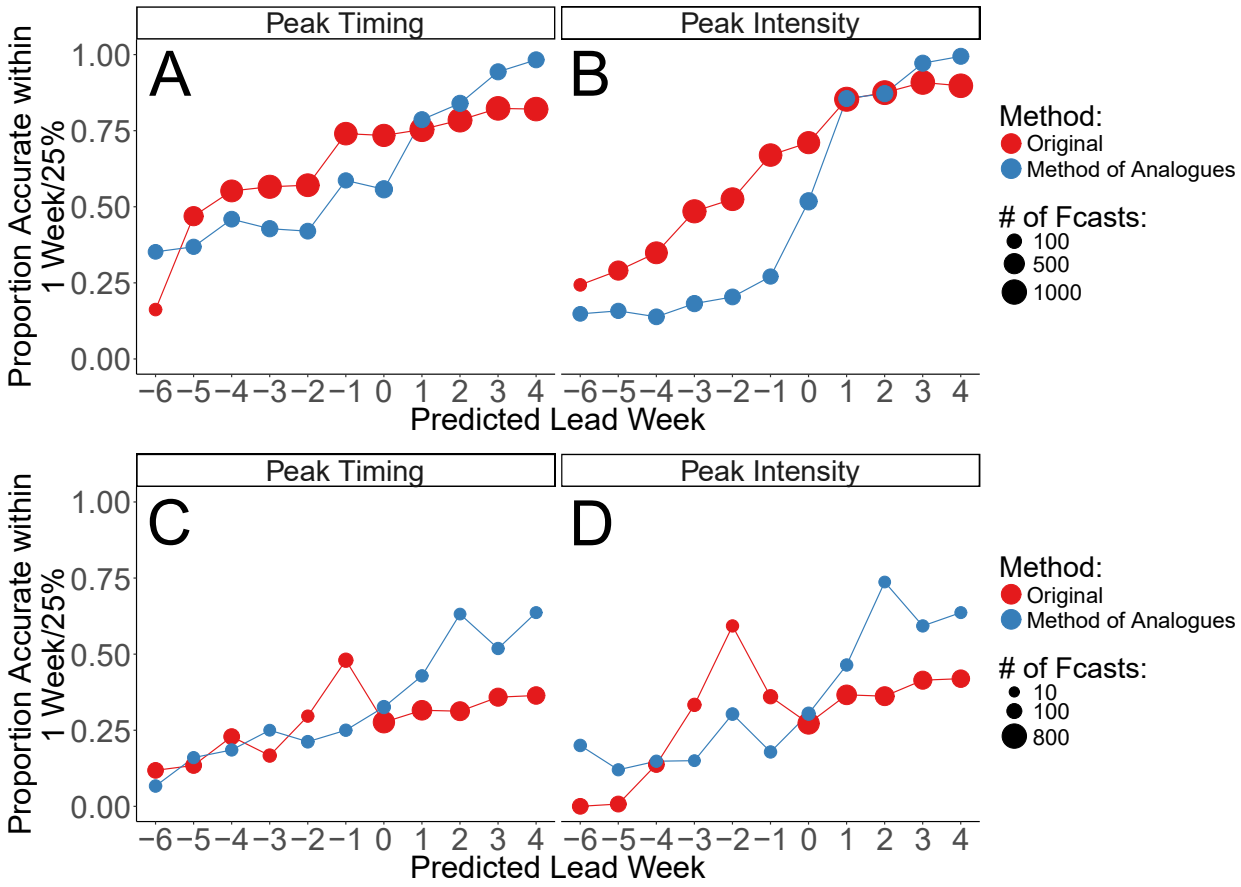
S10 Figure. Histograms of peak timing and intensity forecast error.
Distribution of peak timing (A and C) and intensity (B and D) errors relative to observed for temperate (A and B) and tropical (C and D) regions. To make peak intensity errors comparable between countries, errors are standardized by the observed peak intensity for a given country and season.



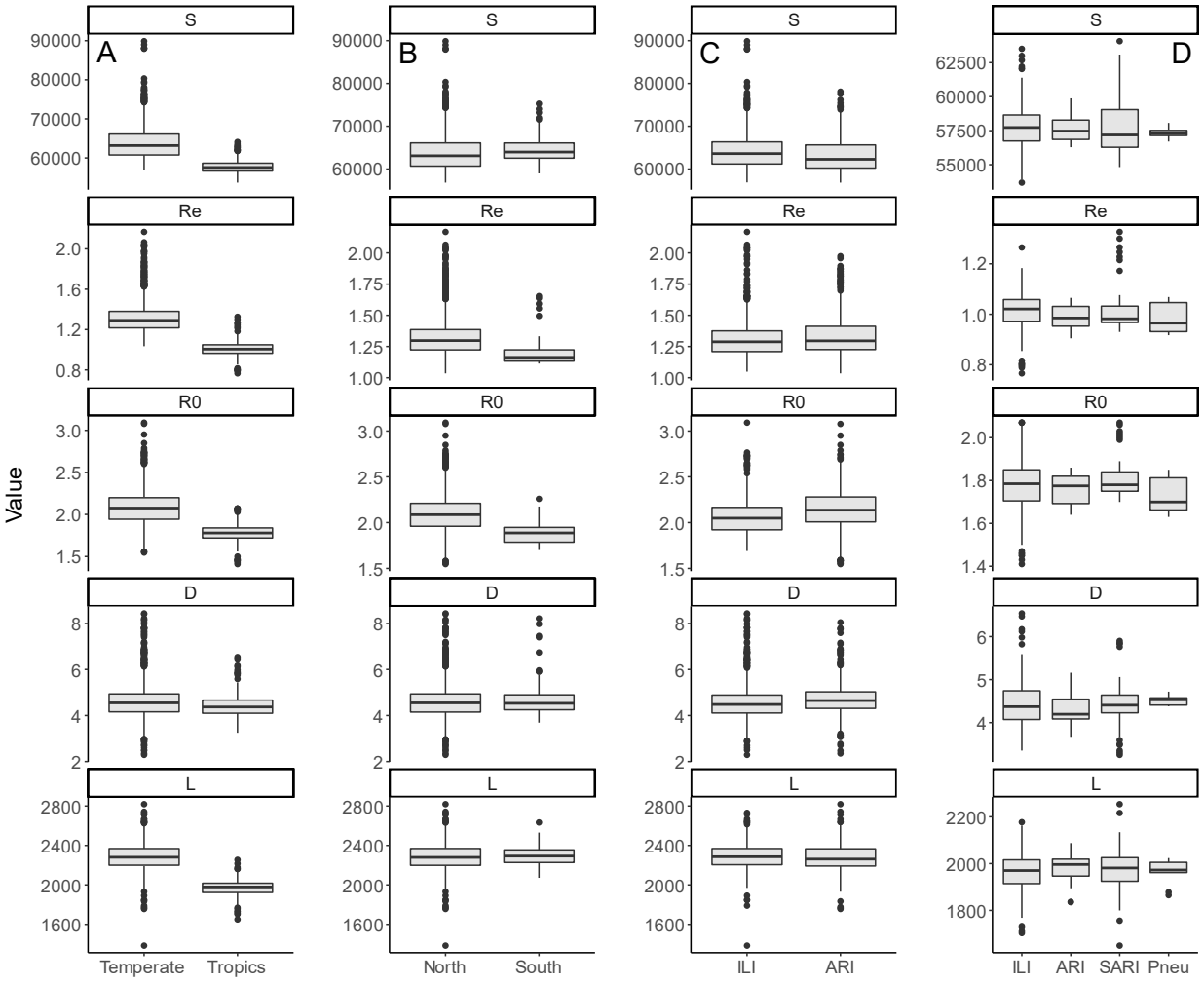
S11 Figure. Forecast accuracy using alternative accuracy cutoffs.
Percent of forecasts accurately predicting peak timing and intensity in temperate (A and C) and tropical (B and D) countries. (A and B) Forecasts are considered accurate if they predict peak timing exactly, and predict peak intensity within 12.5% of the observed value. (C and D) Forecasts are considered accurate when forecasts are within 2 weeks of the observed peak timing and 50% of the observed peak intensity.



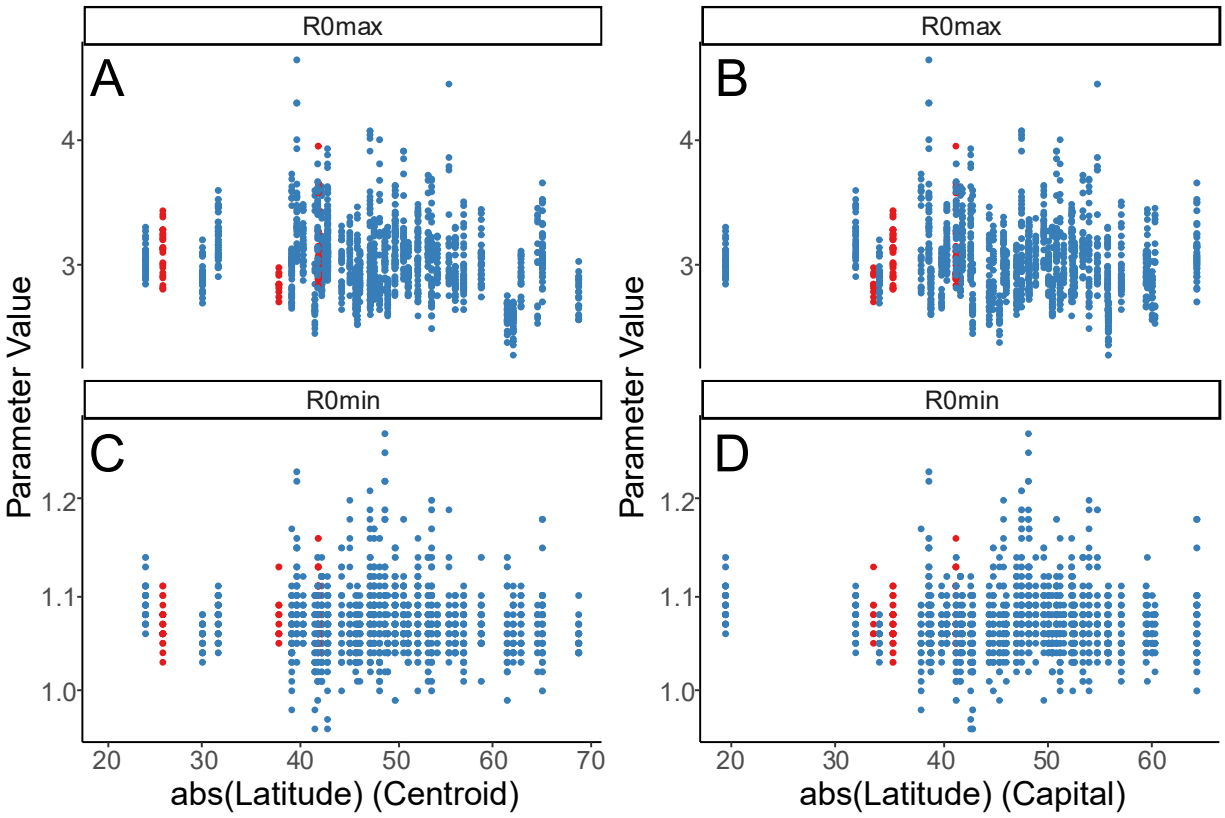
S12 Figure. Forecast accuracy using correlation coefficients and symmetric mean absolute percentage error (sMAPE).
Ranges of correlation coefficients (A and B) and sMAPE (C and D) for temperate (A and C) and tropical (B and D) countries. Points represent median values, and error bars show the 95% credible interval. Point size represents the number of forecasts contributing data to the point in question.



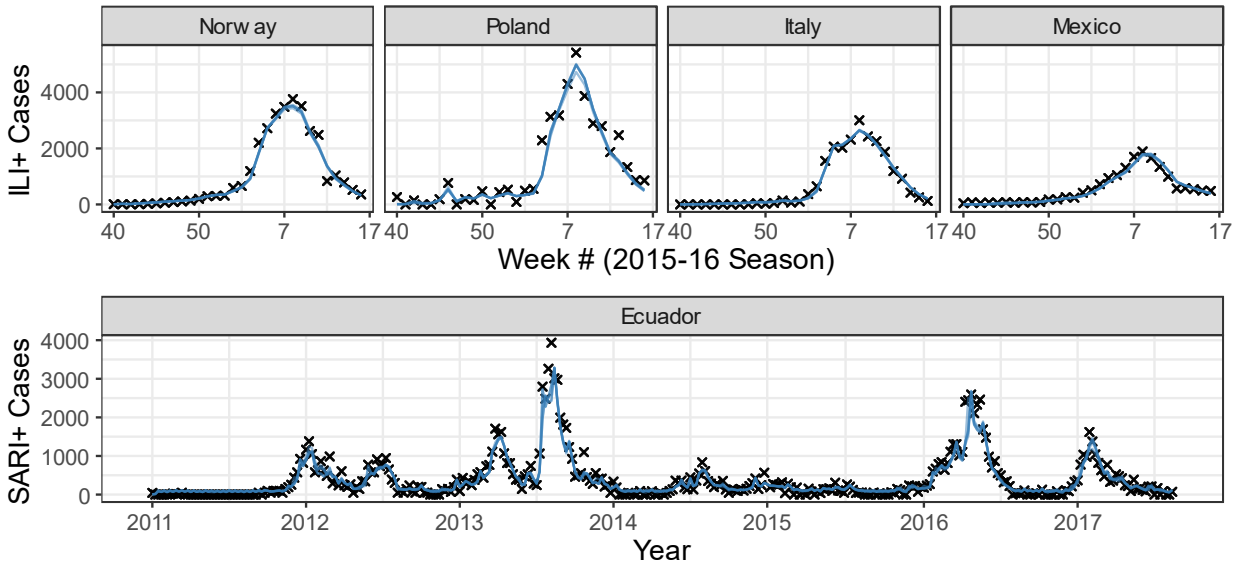
S13 Figure. Forecast accuracy using the method of analogues.
A comparison of peak timing (A and C) and peak intensity (B and D) accuracy in both temperate (A and B) and tropical (C and D) countries between the methods described in the main text (red) and the method of analogues (blue).



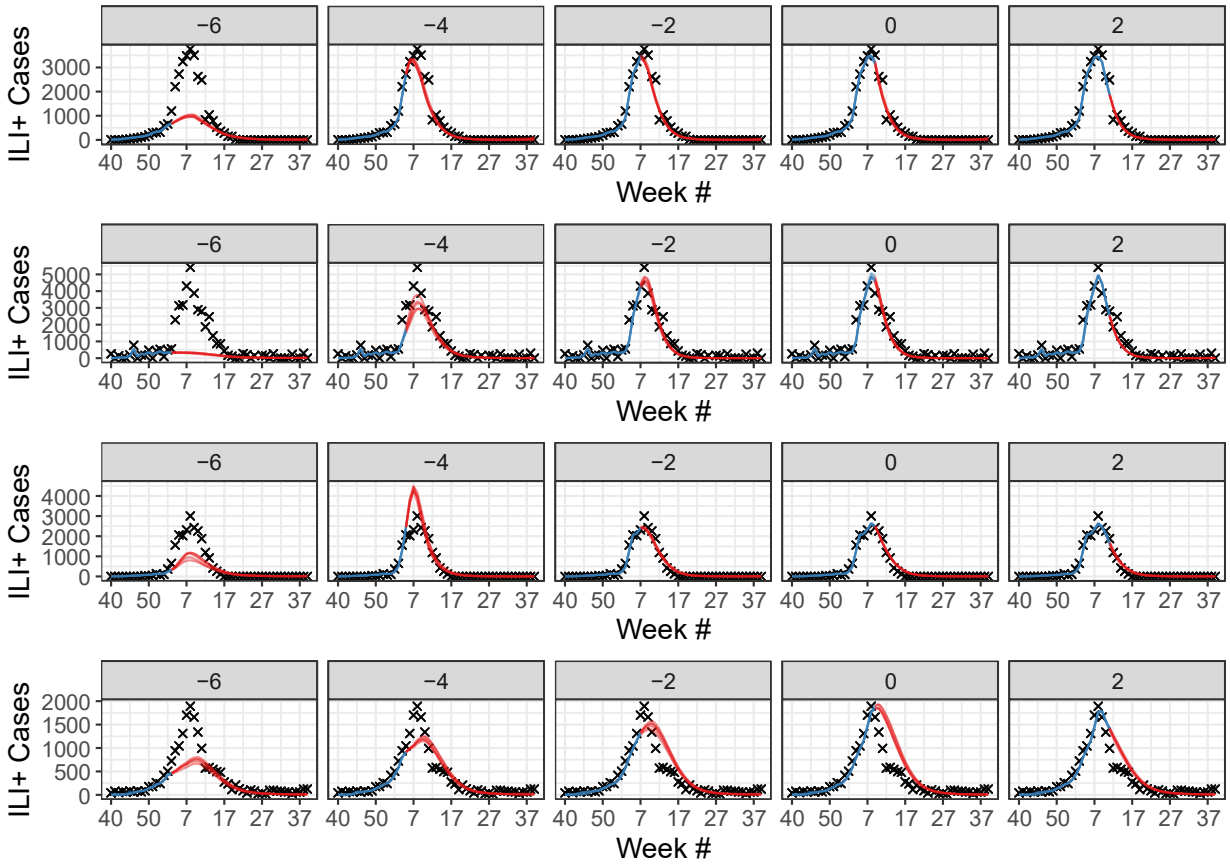
S14 Figure. Inferred model states and parameters.
Ranges for *S0*, R_e, R_0, *D*, and *L* by temperate versus tropics designation (A), hemisphere (B), and data type separated by temperate (C) and tropical (D) regions.



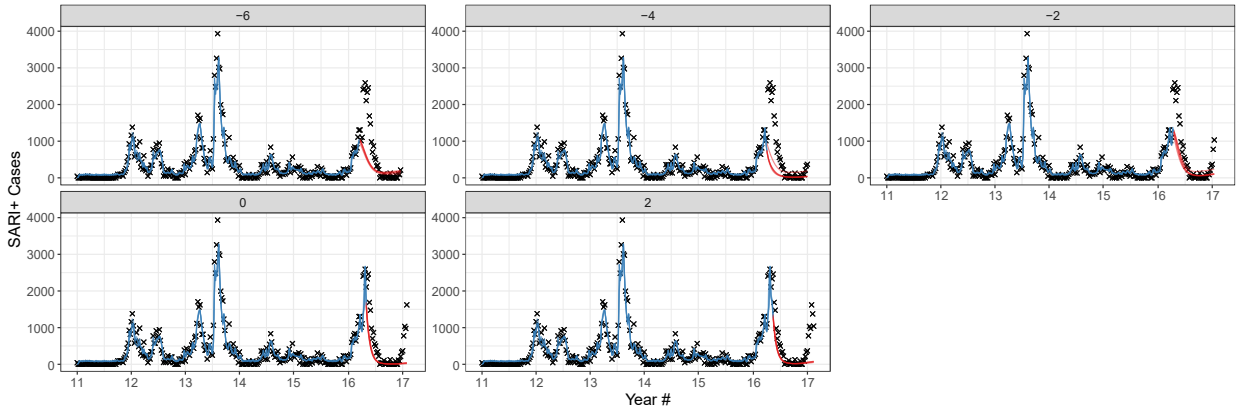
S15 Figure. Ranges of R_{0max} and R_{0min} by latitude.
Distribution of inferred values for R_{0max} (A and B) and R_{0min} (C and D) by latitude (absolute value), defined as the latitude at the country's centroid (A and C) or the latitude of the country's capital (B and D). Values derived from temperate countries are shown in blue, and values from countries in the tropics are in red.



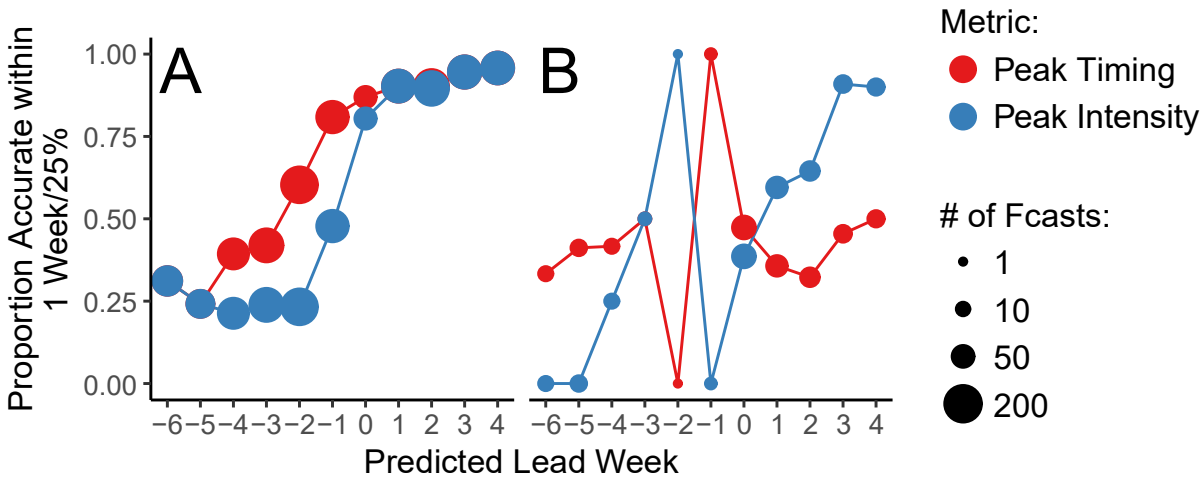
S16 Figure. Posterior visualizations.
Mean posterior fit for 5 models runs of 300 ensemble members each for Norway, Poland, Italy, Mexico, and Ecuador. Fit is plotted for the 2015-16 season for the temperate countries, and for the entire duration of the available data for Ecuador. Observed data are plotted as black x's, while the posterior model fit is plotted in blue.



S17 Figure. Temperate forecast visualizations.
Forecast trajectories for Norway, Poland, Italy, and Mexico for the 2015-16 season. Forecasts are presented starting 6 weeks prior to the observed week through 2 weeks after the peak. Black x's represent observed data, blue lines show model incidence during the training period, and red lines represent forecast trajectory. For each forecast, the 5 runs are shown separately.



S18 Figure. Tropical forecast visualization.
Forecast trajectory for the fifth recorded outbreak in Ecuador in our dataset. Forecasts are presented as in S17 Figure. Because tropical forecasts were generated by fitting the model to the observations continuously, rather than by season, model fitting for all data prior to the fifth outbreak in Ecuador is shown.



S19 Figure. Forecast accuracy for the 2009 influenza pandemic.
Peak timing (red) and intensity (blue) accuracy for forecasts of the 2009 pandemic in temperate (A) and tropical (B) countries.

Chapter 3

Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel

Sarah C Kramer[1], Sen Pei[1], and Jeffrey Shaman[1]

Affiliations:
1. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, 722 W 168th Street, New York, New York, 10032, United States of America

Acknowledgments:
The authors would like to thank Ralf Strobel for helping to improve model efficiency and runtime.

Under Review: Kramer, S.C., Pei, S., & Shaman, J. 2020. Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel. *PLoS Comput Biol.*

Abstract

Past work has shown that models incorporating human travel can improve the quality of influenza forecasts. Here, we develop and validate a metapopulation model of twelve European countries, in which international translocation of virus is driven by observed commuting and air travel flows, and use this model to generate influenza forecasts in conjunction with incidence data from the World Health Organization. We find that, although the metapopulation model fits the data well, it offers no improvement over isolated models in forecast quality. We discuss several potential reasons for these results. In particular, we note the need for data that are more comparable from country to country, and offer suggestions as to how surveillance systems might be improved to achieve this goal.

Author Summary

In our increasingly connected world, infectious diseases are more capable than ever of rapid spread over large geographical distances. Previous research has shown that human travel can be used to better forecast the transmission of influenza, which may in turn help public health and medical practitioners to prepare for outbreaks with increasing lead time. Here, we developed a model of twelve European countries, in which countries are connected based on rates of commuting and air travel between them. We then used this model, along with publicly available influenza incidence data, to forecast future incidence in Europe. We found that forecasts produced with the network model were not more accurate than those produced for individual countries in isolation. We emphasize the need for aligned influenza data collection practices that are comparable between different countries and which will likely improve forecast accuracy.

Introduction

In recent years, multiple studies have demonstrated that skillful influenza forecasts can be generated for a range of cities and countries in both temperate (Biggerstaff et al., 2016; Farrow et al., 2017; Hickmann et al., 2015; Moss et al., 2017; Nsoesie et al., 2014; Shaman et al., 2013; Shaman & Karspeck, 2012; Viboud et al., 2003) and tropical (Ong et al., 2010; Yang et al., 2015) areas. When operationalized, such forecasts can provide early warning for healthcare workers and public health practitioners during both seasonal influenza outbreaks and future influenza pandemics, allowing for a more proactive public health response (*FluSight: Flu Forecasting*, n.d.). Given the high toll of seasonal influenza each year (Iuliano et al., 2018; WHO, n.d.-b), as well as the potential for future pandemic emergence (Herfst et al., 2012; Imai et al., 2012; Taubenberger & Morens, 2006), such proactive responses, informed by real-time forecasts, could reduce influenza-related morbidity and mortality.

Most forecasting to date has been generated for individual locations. City- and country-level outbreaks, however, do not occur in isolation. Research shows that human travel, ranging from long-distance air travel (Brockmann & Helbing, 2013; Brownstein et al., 2006; Crepey & Barthelemy, 2007; Lemey et al., 2014) to short-distance commuting (Balcan et al., 2009; Belik et al., 2011; Bozick & Real, 2015; Charaudeau et al., 2014; Viboud et al., 2003), influences how outbreaks of infectious disease spread. Recent work suggests that, by representing human movement between locations in disease models, forecasts of influenza can be significantly improved at the borough level in New York City (Yang et al., 2016) and at the state level in the United States (Pei et al., 2018). In particular, prediction of local outbreak onset, defined as the first of three consecutive weeks influenza incidence rises above a critical threshold, is difficult to

accurately forecast for a single location in isolation and can be greatly improved by considering commuting and other travel between states (Pei et al., 2018).

Europe represents an interesting opportunity to test the utility of similar models for forecasting international, rather than interstate, influenza transmission. The Schengen Agreement allows citizens and residents of participating countries to cross international borders freely and without border checks, including for employment (*Schengen Area*, n.d.). Thus, although Europe is made up of multiple sovereign countries, commuting and other forms of travel between these countries remain high.

We recently showed that freely available data from the World Health Organization (WHO) can be used to produce skillful forecasts of influenza activity for many European countries in isolation (Kramer & Shaman, 2019). The logical next step, then, is to see whether considering human movement between these countries can yield improvements in forecast accuracy, as was observed at the state level within the United States.

Here we describe the development of a network model of influenza transmission among twelve European countries (Austria, Belgium, Czechia, France, Germany, Hungary, Italy, Luxembourg, the Netherlands, Poland, Slovakia, and Spain), incorporating both cross-border commuting and air travel. We then generate retrospective forecasts of influenza activity using the network model, and compare forecast accuracy to isolated, country-level forecasts. We hypothesize that, by incorporating observed human movement between countries, we will significantly improve forecast accuracy, particularly for onset timing predictions.

Materials and Methods

Influenza Data

Influenza data were obtained from the World Health Organization's (WHO) FluNet (WHO, n.d.-d) and FluID (WHO, n.d.-c) surveillance platforms, which collect virologic and clinical data, respectively, from WHO member countries. Clinical data are reported as numbers of either influenza-like illness (ILI) or acute respiratory infection (ARI), depending on country and season (see Supplementary Materials). Because these measures are based on symptoms that are not specific to influenza, they include cases caused by infections with other respiratory pathogens. To remedy this lack of specificity, we multiplied weekly cases of ILI/ARI by the proportion of tests for influenza that were positive during the same week, as reported in the FluNet data. We refer to this measure as syndromic+ (Kramer & Shaman, 2019).

Although circulating influenza types and (sub)types are often similar between European countries, substantial differences are sometimes observed (see Supplementary Materials). For this reason, we calculated (sub)type-specific syndromic+ cases for each country by multiplying ILI/ARI cases by the proportion of tests positive for H1N1, H3N2, or B influenza. Tests positive for un(sub)typed influenza type A in a given country were assigned as H1N1 or H3N2 in proportion to the respective rates of each (sub)type that week.

For this study, we focus on seasonal influenza. Thus, data were downloaded and processed for the 2010-11 through 2017-18 seasons. Further information on how influenza data were processed can be found in Chapter 2 (Kramer & Shaman, 2019).

Travel Data

Data on both air travel and commuting between European countries were obtained from EuroStat, the European Union's (EU's) statistical office (*Home - Eurostat*, 2018). Specifically, monthly data on the number of passengers carried by aircraft departing from one European country and arriving in another from 2010 through 2017 were obtained from (*Air Passenger Transport*, n.d.), while yearly data on the number of individuals living in one country and working in another for each country pair from 2010 through 2017 were obtained from the Labour Force Survey (*LFS - Overview*, n.d.).

Monthly air travel data were averaged over all available years to yield average travel flows by month, and these data were converted to daily rates. In order to hold model populations constant by country, the travel matrix was made symmetric by averaging the travel rates in both directions along each route.

Unlike the air travel data, commuting data were not available for every possible route. Eurostat calculates country-specific thresholds below which data are either not reported (threshold a) due to low reliability and concern for anonymity, or are reported with a note concerning their reliability (threshold b). We maintained only countries that had at least one incoming and at least one outgoing route with reliable data (i.e. above both thresholds). Because commuting rates along some routes changed substantially over time, we formed "seasonal" commuting matrices by taking the mean of the two yearly matrices contemporary with each influenza season (e.g., for the 2010-11 season, we averaged commuting rates for 2010 and 2011), rather than using an average over all years. Data from 2017 alone were used when forecasting the 2017-18 influenza season. Finally, routes where data were suppressed or unavailable were filled using a random value between 0 and the relevant country-specific threshold a. This was

done separately for each of the 300 ensemble members used during forecasting (see below) to provide a stochastic distribution of possible rates.

Based on the availability of good-quality influenza, air, and commuting data, we included 12 European countries in our model (Figure 1A). More detailed information on data processing can be found in the Supplementary Materials.

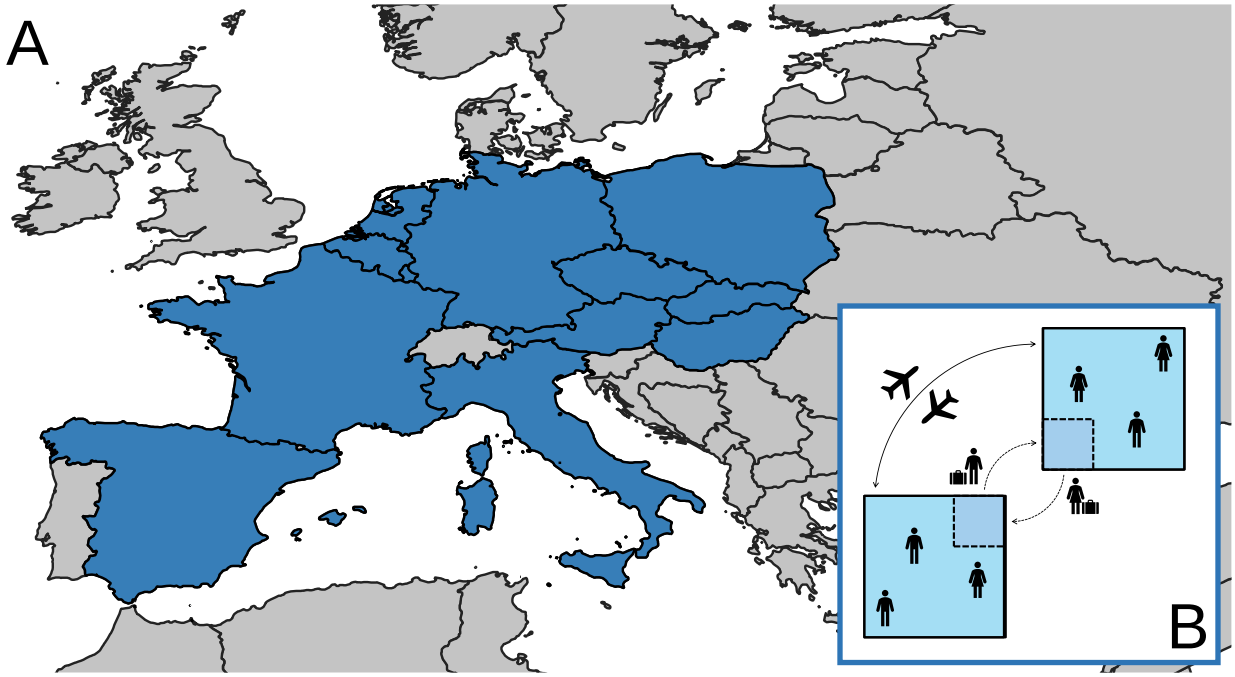


Figure 1. Network model development.
(A) A map of the twelve countries with good-quality influenza and travel data used in the network model. (B) A schematic of a network model with 2 countries and 4 subpopulations. Large squares represent countries, and smaller squares represent residents who commute internationally. The dotted borders on these smaller squares represent their ability to randomly mix with their work country during the day, and their home country at night. Arrows with dotted lines represent asymmetric daily commuting flows, and the solid, double-headed arrow represents symmetric random air travel.

Humidity Data

Absolute humidity data were obtained from NASA's Global Land Data Assimilation System (GLDAS) (Rodell, n.d.). Values were available every three hours and at a spatial resolution of 1°x1° for 1989-2008. We formed 20-year climatologies for each country by aggregating data to the daily level and averaging each daily value over twenty years.

Climatologies were then aggregated to the country level by averaging the climatologies for each individual 1°x1° grid cell within a country, weighted by the number of people living within that grid cell. We weighted by population size in order to better estimate climatic conditions in areas where more people lived, and where more people would therefore be spreading influenza. A more detailed description of these data and how they were processed can be found in Kramer & Shaman (2019).

Network Model

We constructed a networked metapopulation model for the 12 countries, in which subpopulations were assigned to compartments based on both their country of residence and the country in which they work. Within each country, influenza transmission is modeled according to a simple, humidity-forced SIRS model assuming homogenous mixing:

$$\frac{dS}{dt} = \frac{N - S - I}{L} - \frac{\beta(t) I S}{N}$$
$$\frac{dI}{dt} = \frac{\beta(t) I S}{N} - \frac{I}{D} \quad (1)$$

In the above equations, N is the country's population; S and I are the number of people susceptible to and infected with influenza, respectively; t is the time in days; $\beta(t)$ is the rate of transmission at time t; L is the average time before immunity is lost; and D is the mean infectious period. The basic reproductive number, R_0, at time t is equal to $\beta(t)$ times D, and humidity-forcing of transmissibility is modeled by calculating R_0 as:

$$R_0(t) = e^{-180q(t) + \ln(R_{0_{diff}})} + (R_{0_{max}} - R_{0_{diff}}) \quad (2)$$

where R_{0max} is the maximum possible daily basic reproductive number, R_{0diff} is the difference between the maximum and minimum possible basic reproductive number, and $q(t)$ is the absolute humidity on day t (Shaman & Karspeck, 2012). Past work has demonstrated that virus survival and transmissibility between guinea pigs decreases exponentially with increasing absolute humidity (Shaman & Kohn, 2009). However, the exact extent to which this impacts person-to-person transmission remains unknown. R_{0diff} in particular allows modulation of the effect of humidity during a given outbreak. Because the model assumes a single pathogen, we assume that the model parameters L, D, R_{0max}, and R_{0diff} are the same for all countries.

In addition to within-country transmission, individuals are permitted to travel between countries by two methods. First, individuals are assigned to be commuters between given country pairs according to Eurostat data. These individuals spend daytime (eight hours, or one-third of each day) in their designated work country, and nighttime (16 hours) in their country of residence. Commuters mix homogenously with the entire population of the country in which they currently are. Commuting is assumed to be a daily occurrence; thus, commuting in the model captures repeated travel done by the same individuals over time. Because commuting populations are explicitly modeled, we permit the commuting network to remain asymmetric.

Any individual in the model population may also travel to another country randomly, according to observed monthly rates of air travel between European countries. Unlike commuters, "random" travelers are not tracked, and return travel is not explicitly modeled. Furthermore, while commuting is a daily occurrence, random travel is a one-time event. To maintain stable population sizes, daily air travel rates between each country pair are averaged, and this average is used as the rate of travel in both directions. Random travelers may introduce

new infections to a country through their movement, but any propagation of the outbreak occurs solely through the SIRS dynamics.

Because we are primarily interested in the ability of information on international travel to improve forecast accuracy, no random seeding of new infections throughout the outbreak is incorporated. The full equations describing the modeling process can be found in the Supplementary Materials, and a simplified schematic of the model can be seen in Figure 1B.

Retrospective Forecast Generation

We performed retrospective forecasts of (sub)type-specific influenza activity using the network model and influenza data described above in conjunction with a Bayesian data assimilation method. This process consists of a fitting and a forecasting step.

Model Fitting: First, we initiate an ensemble of model runs using random draws from realistic parameter ranges ($2.0 < R_{0max} < 2.8$, $0.2 < R_{0diff} < 1.0$, 2.0 days $< D <$ 7.0 days, 1095 days $< L <$ 3650 days), and integrate the network model forward in time. At each week, we halt the integration and adjust the model state variables (i.e. susceptible, infected, weekly incidence) in all 144 compartments along with the system parameters using the Ensemble Adjustment Kalman Filter (EAKF), a Bayesian data assimilation method commonly used in weather and influenza forecasting (Anderson, 2001; Pei et al., 2018; Shaman & Karspeck, 2012). This adjustment is performed by assimilating observations from each of the twelve countries sequentially through application of Bayes Rule:

$$p(X_t|O_{1:t}) \propto p(X_t|O_{1:(t-1)}) \cdot p(O_t|X_t) \quad (3)$$

where $p(X_t|O_{1:(t-1)})$ is the prior distribution of the modeled cases per 100,000 population in a country given all observations up to but not including time *t*, $p(O_t|X_t)$ is the likelihood of the observed incidence per 100,000 population at time *t* given the modeled incidence rate at time *t*, and $p(X_t|O_{1:t})$ is the posterior distribution of the modeled rate at time *t* given all observations thus far. We note that while the model keeps track of the number of newly infected individuals per compartment, data are available only at the country level. Therefore, modeled incidence is aggregated to the country level prior to the fitting step. Then, the number of susceptible and infected people in each individual compartment, as well as the model parameters, are updated according to cross-ensemble covariability with the country-level rates. Further details on the EAKF and its use in influenza forecasting can be found in the Supplementary Materials and in the literature (Anderson, 2001; Pei et al., 2018; Shaman & Karspeck, 2012; Yang et al., 2014).

Forecasting: After model fitting, forecasts are produced by taking the most recently inferred model states and parameters and running them forward in free simulation until the end of the season. We generate weekly forecasts for each season for weeks 44 through 69, such that forecasts are generated throughout the outbreak. To assess the effects of stochasticity during model initiation, five separate runs, each consisting of 300 ensemble members, are performed.

Choice of Seasons

While extensive cocirculation of influenza (sub)types is common, outbreaks tend to be dominated by one or two (sub)types. To avoid forecasting a given (sub)type during a season where it has little impact, we used a criterion that could also be applied to forecasting in real time. Specifically, forecasting only took place if the positivity rate for a (sub)type of interest

exceeded 10% for three consecutive weeks in at least four of the twelve countries. We therefore forecasted influenza H1N1 for seasons: 2010-11, 2012-13, 2013-14, 2014-15, 2015-16, 2017-18; H3N2 for seasons: 2011-12, 2012-13, 2013-14, 2014-15, 2016-17; and B for seasons: 2010-11, 2012-13, 2014-15, 2015-16, 2016-17, 2017-18.

Choice of Scaling Factors

As described above, model fitting using the EAKF requires that the model output and the data are of the same form. Here, our model output is incidence of a given influenza (sub)type per 100,000 population. Our data, on the other hand, are count data. As most countries do not report reliable or consistent data on the number of patient visits, syndromic+ rates per 100,000 population cannot be directly calculated. However, in past work we have observed that our model systems perform optimally when attack rates fall between 15-50% of a model population size within a season. We therefore calculated a scaling factor, γ, for each country and (sub)type by calculating the range of values that yield attack rates between 15% and 50% for each season, i, ($[\gamma_{15,i}, \gamma_{50,i}]$), then selecting a single value based on the rule:

$$\gamma = \begin{cases} if\ \exists\ \gamma\ \in\ \mathbb{R}: \gamma_{15,i} < \gamma < \gamma_{50,i}\ \forall\ i: & max_{i=1}^{n}(\gamma_{15,i}) \\ else: & min_{i=1}^{n}(\gamma_{50,i}) \end{cases} \quad (4)$$

Thus, scaling factors are allowed to vary by country and (sub)type, but not by season. We consider the scaled data (i.e., γ times syndromic+) to be the estimated syndromic+ rate per 100,000. For each (sub)type, only the seasons listed above in the section "Choice of Seasons" were considered when calculating scaling factors. We previously used this method to calculate scaling factors in our work generating individual country-level forecasts for multiple countries

throughout the world, and found forecasting with the resulting scaled data feasible (Kramer & Shaman, 2019). Scaling values by country and (sub)type are presented in S1 Table.

Individual Country Forecasts

Forecasts for individual countries in isolation were generated as described in Kramer & Shaman (2019). As with the network model, forecasts were generated with 5 to 30 weeks of training data, and the model was fit to each country's data separately using the EAKF. Individual country forecasts used scaling factors as calculated in the previous section. Because no travel between countries is included in the individual models, we seeded new infections in each country at a rate of 0.1 cases per day, or one new case every ten days. Because each country was fit independently, model parameters were allowed to differ by country, unlike in the network model.

Forecast Assessment

We assessed system ability to forecast the timing and magnitude of outbreak peaks (peak timing and intensity, respectively), as well as the timing of outbreak onsets. We define outbreak onset as the first of the first three consecutive weeks to exceed 500 scaled cases, as in our past work using these data (Kramer & Shaman, 2019). For each forecast, we calculated the log score for each of these three metrics by assigning individual ensemble member to bins of size one week (for peak and onset timing) or of size 500 scaled cases (for peak intensity), then calculating the natural logarithm of the proportion of ensemble members falling into the same bin as the observed value. Forecasts for which none of the 300 ensemble members fell within the same bin as the observed value were given a score of -10. The log score has the advantage of being a strictly proper scoring rule, meaning that the expected value has a unique maximum when the

forecast distribution matches the true distribution (Gneiting & Raftery, 2007; Rosenfeld et al., 2018). Additionally, it has been commonly used in past forecasting work, including in the United States Centers for Disease Control and Prevention's (CDC) Predict the Influenza Season Challenge (Biggerstaff et al., 2018).

Because the network model considers information from several countries, it is capable of generating forecasts even at timepoints where a given country has no reported data, unlike an isolated model. To ensure fairness when comparing between the network and isolated models, we considered only those forecasts which are generated by both models.

Forecast Comparison

Although our analysis is not, strictly speaking, paired, we generated forecasts using both the network and isolated models for the same countries, seasons, (sub)types, and weeks. To take advantage of this design, we performed statistical analyses by observed lead week (i.e., the difference between the week of forecast generation and the observed peak or onset week). Specifically, we compared log scores for peak timing, peak intensity, and onset timing between the network and isolated models for forecasts initiated between six weeks prior to and four weeks after the observed peak/onset using the Friedman test (Friedman, 1937). The Friedman test works by ranking the two model forms for each country-season-(sub)type-week pair, then determining whether either model outperforms the other significantly more often. Because we are not interested in forecasts where no onset is predicted, we removed these forecasts from consideration, and country-season-(sub)type-week pairs where either the network or the isolated model (or both) produce no forecasts predicting an onset were removed entirely. Then, because the five model runs for each of these pairs (see Retrospective Forecast Generation above) are not

independent, we randomly chose a single run for each country-season-(sub)type-week before conducting the Friedman test. This process was repeated 100 times, and the median p-value was used.

Results

Influenza Data

Both influenza and travel data were available for a total of 12 countries (Figure 1A). Nine countries had good-quality (Kramer & Shaman, 2019) data for all 8 seasons, two for 7, and one for 6. Outbreaks occurred in between 2 and 12 countries (mean = 9.94, median = 11.0) during each (sub)-type specific seasonal outbreak. Within a given outbreak, the time between the latest and earliest onset ranged from 2 to 12 weeks (mean = 7.65, median = 7.0); differences in peak timing ranged from 2 to 13 weeks (mean = 7.88, median = 8.0). Descriptive statistics broken down by (sub)type can be found in the Supplementary Materials.

Travel Networks

Air travel: In general, rates of air travel were higher in western than eastern Europe. On average, rates of daily air travel were highest in and out of Germany and Spain, while the lowest rates were observed in and out of Slovakia and Luxembourg. The mean daily number of passengers was 2813.57, compared to a median of 741.98, indicating that travel rates were mostly low, with a small number of routes (i.e., country pairs) having particularly high rates. In general, rates of air travel are higher in summer, and lower in winter.

Commuting: For a given season, commuting data between countries were available for between 46 and 58 of 132 possible routes. On average, the number of commuters was by far the highest from France to Luxembourg, and from Poland to Germany. Net commuter inflow was highest for Luxembourg, Germany, and Austria; and net outflow was highest for France, Slovakia, and Poland. Commuting rates generally increased or remained steady over time, with the greatest growth observed among commuters from Slovakia, Czechia, and Spain into Germany. For routes where data in both directions were reported, average commuting rates were highly asymmetric, with travel in one direction on average 12.64 (median = 2.78) times higher than in the other. (As explained above in the Methods, asymmetric commuting flows were permitted in the model.)

Model Fit

In general, the network model was capable of closely fitting the observed data from all countries, despite substantial differences in the intensity of (sub)type activity (Figure 2). Peaks appear to be well-fit in terms of both timing and intensity, although they are occasionally underestimated (see CZ). Fits for all five model runs are very similar, suggesting that the model-inference system is not particularly sensitive to initial conditions. Similar plots for the same season for (sub)types A(H3) and B, found in the supplementary materials (S2 Figure), demonstrate that fit quality remains high, but that the model sometimes has trouble fitting late-season increases in intensity (see S2 Figure B, DE), likely due to filter divergence over time. Over all seasons, model fit by RMSE varied significantly both by (sub)type and country (Kruskal-Wallis tests, $p < 0.00001$). Post-hoc Nemenyi tests revealed that fit quality was significantly highest for A(H1), and lowest for A(H3) ($p < 0.0001$ for all comparisons), and that

fitting for Luxembourg significantly underperformed in comparison to all other countries ($p < 0.0001$ for all comparisons).

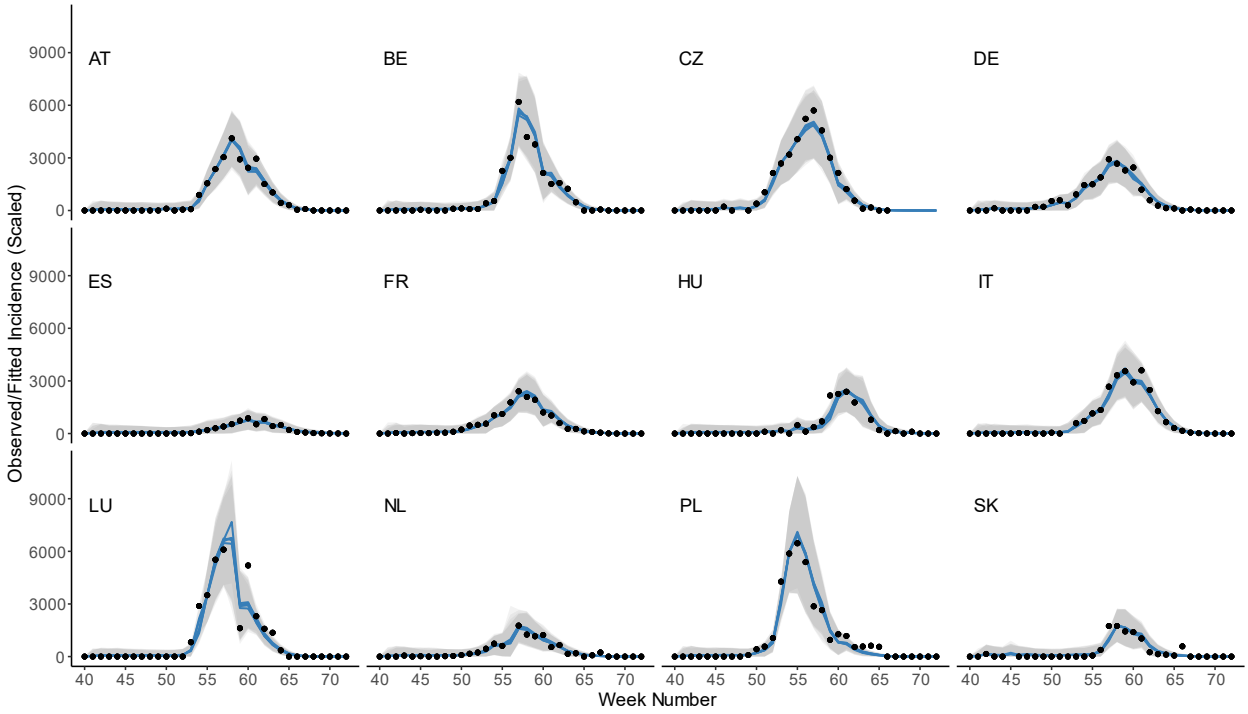


Figure 2. Model fit to observed influenza data.
Scaled observed syndromic+ data for A(H1) throughout the 2012-13 season are shown as points; blue lines represent inferred model incidence. Note that, because each timepoint is fit based on activity in all countries with data, fits are shown for a country even for weeks where that particular country had no available data. Each line represents one of five model runs, each with different starting conditions; 95% confidence intervals (calculated using the ensemble mean, $\bar{X}$, and standard deviation, σ, as $\bar{X} \pm 1.96 \cdot \sigma$) for each run are shown in gray. The root mean square error (RMSE) for this season and (sub)type varied between 60.36 (ES) and 642.99 (LU), and the mean RMSE over all countries was 241.33 (median = 200.09).

Retrospective Forecast Accuracy

The mean log scores for predictions of peak timing, peak intensity, and onset timing by predicted lead week and by (sub)type are shown in Figure 3. Because the timepoint relative to the peak cannot be known in real time, we plot forecast accuracy by predicted lead week, i.e. the difference between the week at which a forecast was initiated and the predicted peak week. Any

forecast not predicting an onset (i.e., where more ensemble members predicted no onset than any single week value) was considered to be unreliable and excluded.

We find that the network model was outperformed by the isolated model for peak timing, especially when the predicted peak was closer (Figure 3A), and performed similarly to the isolated model for peak intensity (Figure 3B). However, we observed a slight improvement in the mean log score for onset timing prior to predicted onset, and especially at predicted leads of -6 to -3 weeks (Figure 3C). Additionally, we note that the network model produced substantially more forecasts than the isolated models at predicted leads -2 and -1, although this pattern does not hold for earlier lead weeks (Table 1). When assessed by observed lead week, the same general patterns emerge (S3 Figure A-C).

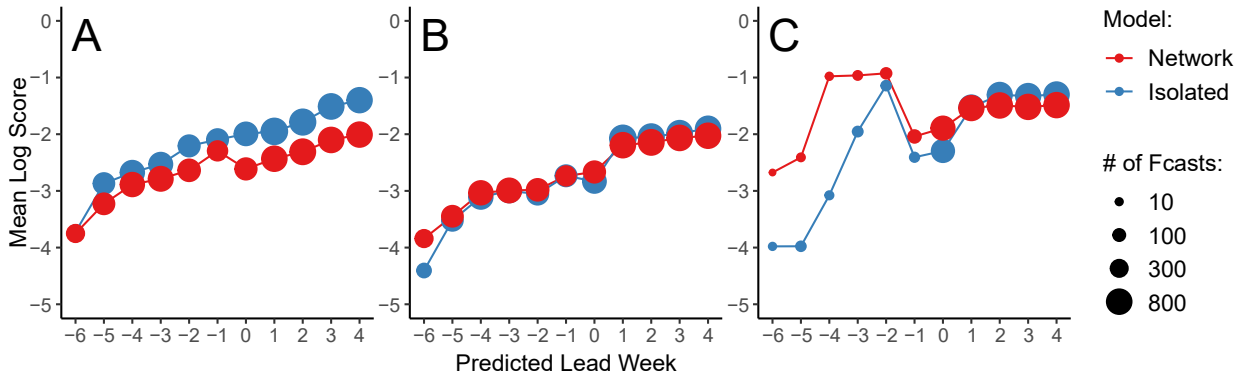


Figure 3. Retrospective forecast accuracy by predicted lead week.
Mean log scores for forecasts generated using the network (red) and isolated (blue) models are shown by predicted lead week for peak timing (A), peak intensity (B), and onset timing (C). The size of the points represents the number of forecasts generated for that lead week for which an onset was predicted.

Table 1. Number of forecasts predicting any onset by predicted onset lead week.

Lead Week	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**
Network	2	17	12	31	63	116	696
Isolated	11	43	20	48	44	37	620

Similar results were observed when forecasts were paired by season, country, subtype, and observed lead week, and only those pairs for which both the network and isolated models predict any onset were maintained (S3 Figure D-F). Friedman tests on these results revealed that the network model performed significantly worse than the isolated model for peak timing ($p < 0.05$) and peak intensity ($p < 0.03$) over all (sub)types and lead weeks combined, although we note that the differences in the log scores between the two models is minimal. When forecasts with log scores of -10 (i.e., forecasts where no ensemble members correctly forecasted the metric in question) were removed, these results were no longer statistically significant ($p > 0.05$). We observed no significant difference in log scores for onset timing predictions ($p > 0.1$).

When results were separated by influenza (sub)type and assessed by predicted lead week, the network model only had more accurate early forecasts of onset timing for A(H1); however, at one week lead, the network model produced more forecasts of onset for all (sub)types (S4 Figure and S2 Table). Friedman tests comparing (sub)type-specific forecasts indicated that the isolated model produced more accurate forecasts of peak timing for (sub)type B only ($p < 0.03$), and forecasts of peak intensity for (sub)type A(H3) only ($p < 0.02$); only the latter remained significant after removing forecasts with log scores equal to -10 ($p < 0.04$). We found no other significant differences between the network and isolated models by (sub)type.

Because so few forecasts are generated before the predicted onset, meaningful comparisons of the results for onset timing by country are not possible. Log scores for peak timing and intensity by country are presented in S5 Figure, and are discussed briefly in the Supplementary Materials.

Forecast Calibration

For a forecast to be useful, it needs to communicate not only a prediction, but also a level of certainty for that prediction. Although the log score considers both sharpness and calibration, we also assessed calibration in isolation by calculating how often observed metrics fell within a range of prediction intervals implied by ensemble spread. As in Figure 3, results are shown by predicted lead week; we focus here only on predictions made before the predicted peak or onset week. Forecasts of peak timing and intensity appear to be well-calibrated for both the network and isolated models, and forecasts of peak intensity using the network model appear to be slightly better-calibrated than those generated in isolation (Figure 4A-B, D-E). The isolated models seem to produce better-calibrated forecasts of onset timing, except at very early leads (Figure 4C and F). However, it is important to note that the sample size here is very small, as not many forecasts were produced prior to outbreak onset (see Table 1).

Discussion

Here we describe the development of a network model for influenza transmission in Europe, and test whether the network model, by incorporating human travel between countries, improves upon isolated country models in forecasting influenza activity. We found that the network model tended to produce more accurate forecasts of onset timing at early predicted leads, and generated substantially more forecasts at predicted leads of -1, indicating that the model can better anticipate when an onset will occur in the following week. However, the number of forecasts that predict any onset timing prior to the predicted onset was few, and the majority of countries and seasons had no or very few forecasts of onset timing until after onset occurs.

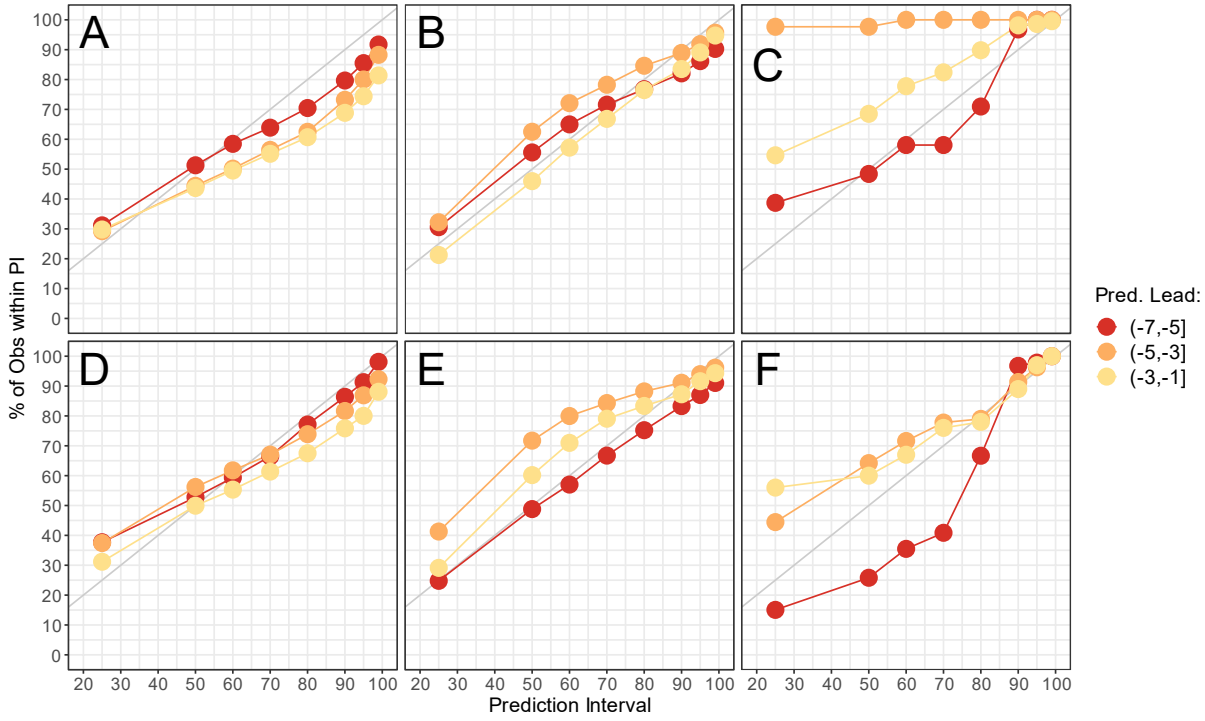


Figure 4. Retrospective forecast calibration for the network (A-C) and isolated (D-F) models. Points show the proportion of observed values for peak timing (A and D), peak intensity (B and E), and onset timing (C and F) that fell within the 25%, 50%, 80%, 90%, 95%, and 99% prediction intervals of 300 ensemble members. Colors represent the predicted lead to the peak (A-B, D-E) or to the onset (C and F). In a perfectly-calibrated model, we expect *n*% of observations to fall within the *n*% prediction interval; this situation is shown as a gray line for reference.

Additionally, we found that the isolated model significantly outperformed the network model in forecasting both peak timing and intensity, although these differences appear to be primarily due to a small number of network forecasts where no ensemble members correctly predict the peak. Furthermore, the magnitude of the differences in log scores was small, and likely not of great practical significance.

Model fit quality was significantly higher for A(H1), and lowest for A(H3), but this did not seem to influence the (sub)type-specific forecasting results. While the network model appeared to improve forecasts of onset timing for A(H1) (S4 Figure), very few forecasts were generated (S2 Table), and the results were not statistically significant. The reason for this discrepancy in fit quality is unclear: all influenza (sub)types generally led to outbreaks in a

similar number of countries, and had similar synchrony with regard to onset and peak timings by country. We also found no evidence of higher signal smoothness for outbreaks of A(H1). However, we did find that scaled outbreaks of A(H3) were significantly larger than outbreaks of the other two (sub)types. It is possible that the network model has more trouble fitting larger peaks, especially when it must simultaneously fit countries where no outbreak onset has occurred. (See Supplementary Materials for descriptive statistics and data quality analyses by (sub)type.)

Overall, we found that despite the inclusion of human travel, the network model did not offer any substantial advantage when it came to influenza forecast generation. These results contrast previous findings, which showed that explicitly modeling travel (in particular commuting) between US states (Pei et al., 2018) and between New York City boroughs (Yang et al., 2016) significantly increased forecast accuracy, especially accuracy for onset timing predictions. While it is not possible to determine the exact roots of this discrepancy, it is likely that differences in data quality play a large role. In particular, we point to data quality issues that hinder comparison between countries. While the influenza data used in this study were likely noisier and more prone to missingness than the US data, we have shown that this is not necessarily associated with lower forecast accuracy in the isolated model (Kramer & Shaman, 2019) or in the network model (see Supplementary Materials). That said, we note here that noisiness in the data from Luxembourg was associated with lower-quality model fit (see “Model Fit” above and Supplementary Materials). It is possible that, while the impact of noisy data may be small on the level of the individual country, the influence of several countries is amplified when all countries must be fit simultaneously.

Likely more important, however, is the requirement that we have some idea of the relative intensity of influenza activity by country in order to properly model influenza transmission between countries. Unfortunately, the syndromic data as reported to FluID rarely have reliable denominator data; countries generally do not report the number of total visitors to sentinel facilities, and even reports of the total number of people within surveillance catchment areas are either missing or inconsistent over time. For this reason, we are unable to calculate rates of ILI or ARI.

Furthermore, although surveillance levels do vary by state within the US, surveillance systems are likely much more variable by country in Europe (WHO & ECDC, 2019). Even if it were possible to calculate rates of syndromic+ cases, heterogeneities and biases in surveillance strategies mean that these country-specific rates may not be comparable to rates from other countries. For example, direct comparison is not possible between countries collecting ILI versus ARI data, or collecting virologic data from all patients versus only those experiencing severe symptoms. These differences, compounded with the lack of denominator data, likely make it difficult to properly simulate and forecast international transmission in Europe. While we attempted to remedy this issue using scaling factors, this approach imperfectly accounts for multiple, possibly shifting, heterogeneities between locations (see Supplementary Materials), and may be insufficient for the model-assimilation system developed here.

The extent to which these data quality issues reflect how the data are collected at the country level, or how the data are reported to the WHO, is unclear. Although a well-established network of laboratories reporting virologic data to the WHO exists (WHO, n.d.-a), the push for centralized collection of syndromic data is recent (Ortiz et al., 2009; WHO, 2014), and reporting of data to FluNet and FluID remains voluntary. It is likely that many countries do not report all

of the data they collect. However, obtaining and formatting data from several countries independently is time-consuming and, especially in the case of an emerging pandemic, impractical. Thus, barriers to timely and skillful forecasts of influenza transmission throughout Europe exist due to how data are collected, as well as how and whether data are reported to central databases.

In order to improve model performance, we recommend that, at minimum, denominator data for sentinel surveillance efforts be reported, preferably as the number of total visits (not influenza-specific visits) made to sentinel sites. We also recommend that, where possible, countries use ILI rather than ARI, as it is more specific, and we have previously shown that ILI data yield more skillful forecasts of influenza activity (Kramer & Shaman, 2019). While it would of course be helpful if surveillance systems were more similar country to country, we recognize that the ideal surveillance strategy for a country will depend on factors such as country size, healthcare system, and goals related to surveillance. The WHO does offer extensive guidance on improving influenza surveillance systems, particularly syndromic surveillance systems, which historically have been less developed than virologic surveillance systems, and are likely responsible for the majority of data quality issues encountered in this work (WHO, 2014). We note there was a significant push to improve syndromic surveillance and reporting in the wake of the 2009 influenza pandemic (Ortiz et al., 2009; WHO, 2010). Thus, we expect that improvement is indeed possible, but may depend on the extent to which influenza is seen as a public health priority.

We also note that the commuting data used here are likely of lower quality than those used in the United States. In the US, commuting data are captured by government census. While the European Labour Force Survey uses standardized definitions and questionnaires in order to

bolster comparability between countries, data collection is still the responsibility of individual countries, meaning that data collection is inherently less centralized and standardized than in the US (European Commission, 2003). Furthermore, as we note in the Methods above, data below certain thresholds are not reported to users at all. However, because commuting along these routes is minimal, we do not expect this missingness to have greatly impacted results. Overall, we expect that the network model's inability to improve forecast accuracy is primarily driven by low-quality influenza, and not travel, data.

It is also possible that the model itself is not properly specified for exploring influenza transmission in Europe. In other words, while a model incorporating commuting data between locations may be appropriate for interstate influenza transmission in the US, commuting and air travel may not be as important at driving cross-border transmission in Europe. Despite high rates of cross-border commuting at border regions, international commuting in Europe still only accounts for 0.9% of commuting overall (*Commuting Patterns*, n.d.). Influenza transmission may therefore rely more on non-commuting train and automobile traffic, neither of which are captured in our model. More robust data on train travel, and especially on automobile travel across country borders, could contribute to a better specified model, able to anticipate outbreak onsets based on activity in countries heavily linked by train routes and roads. Alternatively, commuting may be an important driver of influenza transmission in border regions, but country-level influenza overall may be much more heavily driven by travel patterns within the country itself. Models focusing on travel on this smaller sub-national scale may therefore have more success forecasting country-level influenza transmission.

Finally, we note that, as with any model, we are unable to capture all features of the system we seek to model. R_0 is a composite parameter that takes into account the intrinsic

transmissibility of a given pathogen, as well extrinsic factors driving transmission, such as environmental factors or contact networks (Keeling & Eames, 2005; Lipsitch, 2003). Our network model only allows differences in R_0 by country on the basis of country-level absolute humidity. However, we recognize that contact patterns may be driven by population density, age structure, school holiday schedules, and various other demographic and cultural factors not captured here that nonetheless can vary greatly by country (Bansal et al., 2007; Eames et al., 2012; Ferrari et al., 2011). While it may be possible to implicitly consider these differences in our model by allowing R_{0max} and R_{0diff} to vary by country, explicit consideration of these differences is problematic because we do not understand exactly how they may influence influenza transmission. While it can be tempting to add complexity to a model in an effort to increase its fit of observations, adding details without understanding their real-life impact has the potential to backfire, leading to overfit results that are relevant only to the model population and not to reality (Naimi, 2016; Sterman, 2006), and which may also corrupt forecast accuracy. Future work should attempt to better understand the drivers of international influenza spread not just between European countries, but on a variety of spatial scales, both smaller and larger.

Conclusions

Here we present a novel network model of influenza transmission among twelve European countries, and test the model ability to improve influenza forecasting accuracy over isolated, country-level models. While the network model system was unable to improve forecasts in most circumstances, the success of similar network models in the United States suggests that this could be a powerful tool for improving influenza forecasts if data quality issues were to be addressed. We identify key opportunities for improvement in data collection and sharing that

may allow for success in the future. In the meantime, future work should focus on better understanding the various drivers of international influenza transmission in Europe and globally, so that models can better account for these relevant factors.

References

Air passenger transport between countries. (n.d.). Eurostat. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=avia_paocc&lang=en

Anderson, J. L. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, *106*(51), 21484–21489.

Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When individual behaviour matters: Homogeneous and network models in epidemiology. *Journal of The Royal Society Interface*, *4*(16), 879–891. https://doi.org/10.1098/rsif.2007.1100

Belik, V., Geisel, T., & Brockmann, D. (2011). Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases. *Physical Review X*, *1*(1), 011001. https://doi.org/10.1103/PhysRevX.1.011001

Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I. C.-H., Hickmann, K. S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., Velardi, P., Vespignani, A., & Finelli, L. (2016). Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infectious Diseases*, *16*. https://doi.org/10.1186/s12879-016-1669-x

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. (2018). Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*. https://doi.org/10.1016/j.epidem.2018.02.003

Bozick, B. A., & Real, L. A. (2015). The Role of Human Transportation Networks in Mediating the Genetic Structure of Seasonal Influenza in the United States. *PLOS Pathogens*, *11*(6), e1004898. https://doi.org/10.1371/journal.ppat.1004898

Brockmann, D., & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science*, *342*(6164), 1337–1342.

Brownstein, J. S., Wolfe, C. J., & Mandl, K. D. (2006). Empirical Evidence for the Effect of Airline Travel on Inter-Regional Influenza Spread in the United States. *PLoS Medicine*, *3*(10), e401. https://doi.org/10.1371/journal.pmed.0030401

Charaudeau, S., Pakdaman, K., & Boëlle, P.-Y. (2014). Commuter Mobility and the Spread of Infectious Diseases: Application to Influenza in France. *PLoS ONE*, *9*(1), e83002. https://doi.org/10.1371/journal.pone.0083002

Crepey, P., & Barthelemy, M. (2007). Detecting Robust Patterns in the Spread of Epidemics: A Case Study of Influenza in the United States and France. *American Journal of Epidemiology*, *166*(11), 1244–1251. https://doi.org/10.1093/aje/kwm266

Eames, K. T. D., Tilston, N. L., Brooks-Pollock, E., & Edmunds, W. J. (2012). Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. *PLoS Computational Biology*, *8*(3), e1002425. https://doi.org/10.1371/journal.pcbi.1002425

Employment and unemployment (LFS)—Overview. (n.d.). Eurostat. https://ec.europa.eu/eurostat/web/lfs

European Commission. (2003). *The European Union labour force survey—Methods and definitions—2001*. European Communities. https://unstats.un.org/unsd/EconStatKB/Attachment269.aspx?AttachmentType=1

Farrow, D. C., Brooks, L. C., Hyun, S., Tibshirani, R. J., Burke, D. S., & Rosenfeld, R. (2017). A human judgment approach to epidemiological forecasting. *PLoS Computational Biology*, *13*(3). https://doi.org/10.1371/journal.pcbi.1005248

Ferrari, M. J., Perkins, S. E., Pomeroy, L. W., & Bjørnstad, O. N. (2011). Pathogens, Social Networks, and the Paradox of Transmission Scaling. *Interdisciplinary Perspectives on Infectious Diseases*, *2011*, 1–10. https://doi.org/10.1155/2011/267049

FluSight: Flu Forecasting. (n.d.). Centers for Disease Control and Prevention. https://www.cdc.gov/flu/weekly/flusight/index.html

Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, *32*(200), 675–701. https://doi.org/10.1080/01621459.1937.10503522

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Herfst, S., Schrauwen, E. J. A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V. J., Sorrell, E. M., Bestebroer, T. M., Burke, D. F., Smith, D. J., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., & Fouchier, R. A. M. (2012). Airborne transmission of influenza A/H5N1 virus between ferrets. *Science (New York, N.Y.)*, *336*(6088), 1534–1541. https://doi.org/10.1126/science.1213362

Hickmann, K. S., Fairchild, G., Priedhorsky, R., Generous, N., Hyman, J. M., Deshpande, A., & Del Valle, S. Y. (2015). Forecasting the 2013–2014 Influenza Season Using Wikipedia. *PLoS Computational Biology*, *11*(5). https://doi.org/10.1371/journal.pcbi.1004239

Home—Eurostat. (2018, July 25). Eurostat. http://ec.europa.eu/eurostat/web/main/home

Imai, M., Watanabe, T., Hatta, M., Das, S. C., Ozawa, M., Shinya, K., Zhong, G., Hanson, A., Katsura, H., Watanabe, S., Li, C., Kawakami, E., Yamada, S., Kiso, M., Suzuki, Y.,

Maher, E. A., Neumann, G., & Kawaoka, Y. (2012). Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature*, *486*(7403), 420–428. https://doi.org/10.1038/nature10831

Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., Cohen, C., Gran, J. M., Schanzer, D., Cowling, B. J., Wu, P., Kyncl, J., Ang, L. W., Park, M., Redlberger-Fritz, M., Yu, H., Espenhain, L., Krishnan, A., Emukule, G., … Mustaquim, D. (2018). Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *The Lancet*, *391*(10127), 1285–1300. https://doi.org/10.1016/S0140-6736(17)33293-2

Keeling, M. J., & Eames, K. T. D. (2005). Networks and epidemic models. *Journal of The Royal Society Interface*, *2*(4), 295–307. https://doi.org/10.1098/rsif.2005.0051

Kramer, S. C., & Shaman, J. (2019). Development and validation of influenza forecasting for 64 temperate and tropical countries. *PLOS Computational Biology*, *15*(2), e1006742. https://doi.org/10.1371/journal.pcbi.1006742

Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D., & Suchard, M. A. (2014). Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathogens*, *10*(2), e1003932. https://doi.org/10.1371/journal.ppat.1003932

Lipsitch, M. (2003). Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science*, *300*(5627), 1966–1970. https://doi.org/10.1126/science.1086616

Moss, R., Zarebski, A., Dawson, P., & McCAW, J. M. (2017). Retrospective forecasting of the 2010–2014 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology & Infection*, *145*(1), 156–169. https://doi.org/10.1017/S0950268816002053

Naimi, A. I. (2016). Commentary: Integrating Complex Systems Thinking into Epidemiologic Research. *Epidemiology*, *27*(6), 843–847. https://doi.org/10.1097/EDE.0000000000000538

Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., & Marathe, M. V. (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses*, *8*(3), 309–316. https://doi.org/10.1111/irv.12226

Ong, J. B. S., Chen, M. I.-C., Cook, A. R., Lee, H. C., Lee, V. J., Lin, R. T. P., Tambyah, P. A., & Goh, L. G. (2010). Real-Time Epidemic Monitoring and Forecasting of H1N1-2009 Using Influenza-Like Illness from General Practice and Family Doctor Clinics in Singapore. *PLoS ONE*, *5*(4), e10036. https://doi.org/10.1371/journal.pone.0010036

Ortiz, J. R., Sotomayor, V., Uez, O. C., Oliva, O., Bettels, D., McCarron, M., Bresee, J. S., & Mounts, A. W. (2009). Strategy to Enhance Influenza Surveillance Worldwide. *Emerging Infectious Diseases*, *15*(8), 1271–1278. https://doi.org/10.3201/eid1508.081422

Pei, S., Kandula, S., Yang, W., & Shaman, J. (2018). Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences*, 201708856. https://doi.org/10.1073/pnas.1708856115

Rodell, M. (n.d.). *LDAS | Land Data Assimilation Systems* [Text.Journal]. https://ldas.gsfc.nasa.gov/gldas/GLDASgoals.php

Rosenfeld, R., Grefenstette, J., & Burke, D. (2018). *A Proposal for Standardized Evaluation of Epidemiological Models*. https://delphi.midas.cs.cmu.edu/files/StandardizedEvaluation_Revised_12-11-09.pdf

Schengen Area. (n.d.). https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/schengen_en

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, *109*(50), 20425–20430. https://doi.org/10.1073/pnas.1208772109

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3837

Shaman, J., & Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3243–3248. https://doi.org/10.1073/pnas.0806852106

Statistics on commuting patterns at regional level—Statistics Explained. (n.d.). Eurostat. https://ec.europa.eu/eurostat/statistics-explained/pdfscache/50943.pdf

Sterman, J. D. (2006). Learning from Evidence in a Complex World. *American Journal of Public Health*, *96*(3), 505–514. https://doi.org/10.2105/AJPH.2005.066043

Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: The mother of all pandemics. *Rev Biomed*, *17*, 69–79.

Viboud, C., Boëlle, P.-Y., Carrat, F., Valleron, A.-J., & Flahault, A. (2003). Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, *158*(10), 996–1006.

WHO. (n.d.-a). *Global Influenza Surveillance and Response System (GISRS)*. World Health Organization. https://www.who.int/influenza/gisrs_laboratory/en/

WHO. (n.d.-b). *Influenza (Seasonal)*. WHO. https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)

WHO. (n.d.-c). *WHO | FluID - a global influenza epidemiological data sharing platform*. WHO. https://www.who.int/influenza/surveillance_monitoring/fluid/en/

WHO. (n.d.-d). *WHO | FluNet*. WHO. https://www.who.int/influenza/gisrs_laboratory/flunet/en/

WHO. (2010). *Surveillance Recommendations for Member States in the Post Pandemic Period*. World Health Organization. https://www.who.int/csr/resources/publications/swineflu/surveillance_post_pandemic_20100812/en/

WHO. (2014). *Global Epidemiological Surveillance Standards for Influenza*. World Health Organization. https://www.who.int/influenza/resources/documents/influenza_surveillance_manual/en/

WHO, & ECDC. (2019). *Influenza Surveillance: Country, Territory and Area Profiles 2019*. World Health Organization Regional Office for Europe. http://www.euro.who.int/__data/assets/pdf_file/0016/402082/InfluenzaSurveillanceProfiles_2019_en.pdf

Yang, W., Cowling, B. J., Lau, E. H. Y., & Shaman, J. (2015). Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology*, *11*(7), e1004383. https://doi.org/10.1371/journal.pcbi.1004383

Yang, W., Karspeck, A., & Shaman, J. (2014). Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLoS Computational Biology*, *10*(4), e1003583. https://doi.org/10.1371/journal.pcbi.1003583

Yang, W., Olson, D. R., & Shaman, J. (2016). Forecasting Influenza Outbreaks in Boroughs and Neighborhoods of New York City. *PLoS Computational Biology*, *12*(11), e1005201.

S1 Text. Supplementary methods and results.

Supplementary Methods

Influenza Data Processing

The syndromic data used for this study were ILI data except in Germany and Luxembourg, which preferentially reported ARI data, as well as France, which reported ARI data during the 2012-13 and 2013-14 seasons. No virologic (FluNet) data were available for France for the 2010-11 and 2011-12 seasons. Additionally, data from Czechia in 2013-14 and Poland in 2011-12 were removed from consideration because the attack rates of total syndromic+ (including all (sub)types) for these outbreaks were less than 5% of the attack rate of the largest outbreak in these countries, an exclusion criteria laid out in Kramer & Shaman (2019)).

Travel Data Processing

Air Data: Because air travel data are collected from both source and destination countries, the number of passengers traveling each route is, theoretically, reported twice. Due to close agreement between the data reported by source and destination countries, we chose to simply use the data as reported by source countries. Data were averaged over all years in order to preserve data along routes where data were available some years but not others. Because travel was assumed to be symmetric, if data along a route were missing in one direction but not the other, the available value was assumed to be the travel rate in both directions.

Commuting Data: The European Union (EU) Labour Force Survey is a survey of private households in 35 countries, including all EU member states. Data collection is the responsibility

of individual countries, so while questions and definitions remain constant across all countries, sampling strategies may not, and results are therefore not perfectly comparable between countries (*LFS - Main Features*, n.d.; European Commission, 2003). Of the routes where no data were reported, all but four were unreported due to being below threshold 'a' (see main text) and were not simply missing. We note that these routes tended to be between countries that were geographically distant, or else represented routes where economic incentive to commute was low (Mathä & Wintr, 2009; Decoville et al., 2013). Finally, although Ireland had reliable incoming and outgoing commuting data, the United Kingdom did not. Removing the United Kingdom from the model left Ireland isolated geographically, and we therefore subsequently removed Ireland from consideration as well.

Type and Subtype Dynamics in Europe

For each season and country, we calculated the percentage of positive influenza tests in the FluID data that were H1N1, H3N2, un(sub)typed A, and B. After allocating un(sub)typed influenza A proportionally to H1N1 or H3N2, the resulting (sub)type-specific, scaled syndromic+ data for each country over the study period can be seen in S1 Figure. Although the proportional contribution of each (sub)type by season is broadly similar across countries, substantial differences do occur. For example, during the 2014-15 season, A(H1) was the dominant (sub)type in Italy (although A(H3) was also present), while A(H3) dominated in most other countries. In the 2015-16 season, A(H1) and B circulated in all countries, but the relative activity of the two (sub)types varied by country.

Network Model

The full equations for the network model during daytime are:

$$\frac{dS_n^k}{dt} = \frac{N_n^k - S_n^k - I_n^k}{L} - \frac{\beta_n(t) S_n^k I_n}{N_n} - \frac{S_n^k}{N_n}\sum_{m \neq n} r(n,m) + \frac{N_n^k}{N_n}\sum_{m \neq n}\left[r(m,n)\sum_h \frac{S_m^h}{N_m}\right]$$

$$\frac{dI_n^k}{dt} = \frac{\beta_n(t) S_n^k I_n}{N_n} - \frac{I_n^k}{D} - \frac{I_n^k}{N_n}\sum_{m \neq n} r(n,m) + \frac{N_n^k}{N_n}\sum_{m \neq n}\left[r(m,n)\sum_h \frac{I_m^h}{N_m}\right]$$

where any model state X_n^k is the number of susceptible or infected people who live in country *k* and work in country *n*, X_n is the number of susceptible or infected people currently in location *n*, β_n is the transmission rate in country *n*, and $r(n,m)$ is the daily rate of air travel from country *n* to country *m*; all other parameters are as described in the main text. During nighttime, the equations are:

$$\frac{dS_n^k}{dt} = \frac{N_n^k - S_n^k - I_n^k}{L} - \frac{\beta_k(t) S_n^k I^k}{N^k} - \frac{S_n^k}{N^k}\sum_{h \neq k} r(k,h) + \frac{N_n^k}{N^k}\sum_{h \neq k}\left[r(h,k)\sum_m \frac{S_m^h}{N^h}\right]$$

$$\frac{dI_n^k}{dt} = \frac{\beta_k(t) S_n^k I^k}{N^k} - \frac{I_n^k}{D} - \frac{I_n^k}{N^k}\sum_{h \neq k} r(k,h) + \frac{N_n^k}{N^k}\sum_{h \neq k}\left[r(h,k)\sum_m \frac{I_m^h}{N^h}\right]$$

where X^k is the number of susceptible or infected people currently in location *k*, and the other states and parameters are as described before. Because individuals are assumed to spend 8 hours in the country where they work, and 16 in their home country, daytime equations are multiplied by 1/3, and nighttime equations by 2/3.

Scaling Factors

The use of scaling factors in influenza forecasting is rooted in Bayes' Rule, such that:

$$p(i) = \frac{p(m)}{p(m|i)} p(i|m) = \gamma \cdot (syndromic+)$$

where *p(i)* is the probability of influenza infection (the quantity estimated by our model), *p(m)* is the probability that one seeks healthcare for any reason, *p(m|i)* is the probability of seeking healthcare among those with an influenza infection, and *p(i|m)* is the probability that one is infected with influenza given that one sought healthcare. This final quantity is analogous to our syndromic+ plus measure, which estimates the number of influenza cases at a given time based on data collected from those seeking healthcare for ILI or ARI. The scaling factor, *γ,* therefore represents the probability of seeking medical attention for any reason, divided by the probability of seeking medical attention conditional on infection with influenza, quantities that may be expected to vary based on surveillance systems, disease severity, and health seeking behavior (Morita et al., 2018; Shaman et al., 2013). By multiplying syndromic+ data by such scaling factors, we are theoretically calculating the probability of influenza infection, which is equivalent to the output of our models, and allows the appropriate use of the EAKF.

However, as explained in the main text, the WHO data are typically reported as counts, not rates. Thus, our scaling factors must also account for differences in country population sizes and the size of surveillance catchment areas, and we expect our scaling factors to vary considerably by country. Indeed scaling factors range from 0.033 (France, A(H3)) to 83.24 (Luxembourg, A(H1)) (S1 Table).

Ensemble Adjustment Kalman Filter

Observational Error Variance: Use of the EAKF requires that the degree of error in both the simulated model output and in the data are specified. While the model error can easily be

calculated as the variance of the 300 ensemble members, the error in our observations is unknown. We specify the observational error variance at time *t* as:

$$OEV_t = b + \frac{\left(\sum_{j=t-2}^{t} \frac{O_t}{3}\right)^2}{c}$$

where O_t is the observed syndromic+ data at time *t*. Here, the parameters *b* and *c* were set to 1e5 and 10, respectively, based on a preliminary grid search, but can be altered for different data sources (Morita et al., 2018).

Filter Divergence: As the outbreak progresses and successive data point are fit, there is a tendency for the variance between the ensemble members to shrink, potentially leading to filter divergence, in which the ensemble error variance becomes so low that data points are essentially no longer considered in the fitting process. We attempted to prevent divergence by multiplicatively inflating the prior model variance by 1.05 at each time step, prior to data assimilation (Pei et al., 2018; Shaman et al., 2013).

Supplemental Results

Descriptive Statistics by (Sub)type

The number of countries where outbreaks occurred during a (sub)type-specific seasonal influenza outbreak ranged from 8 to 12 countries (mean = 10.8 median = 11.5) for A(H1), from 8 to 12 countries (mean = 9.6, median = 10.0) for A(H3), and from 2 to 12 countries (mean = 9.3, median = 11.5) for B. Therefore, it appears that the number of countries experiencing outbreaks did not differ notably by subtype, with the exception of a season where at least 4 countries

reported over 10% positivity rates for influenza B, but only two countries experienced any outbreak onset as determined using scaled syndromic+ data. We note that this outbreak of B influenza was particularly late in the season (onsets at weeks 56 and 58), and that the number of clinical cases was already quite low by this point in the outbreak, which explains the lack of onsets despite positivity rates being above our threshold of 10%.

Within an outbreak, the time between the earliest and latest outbreak onsets ranged from 5 to 11 weeks (mean = 7.3, median = 6.5) for A(H1), 5 to 10 weeks (mean = 8, median = 8) for A(H3), and 2 to 12 weeks (mean = 7.7, median = 7) for B; the time between the earliest and latest outbreak peaks was between 6 and 13 weeks (mean = 8.3, median = 7) for A(H1), 5 and 11 weeks (mean = 8, median = 8) for A(H3), and 2 and 12 weeks (mean = 7.3, median = 8) for B. This suggests that, by this rough metric, outbreak synchrony between countries did not vary substantially by subtype. Finally, we find that peak timing (Kruskal-Wallis test, $p < 0.005$), but not onset timing ($p > 0.1$), differed significantly by subtype over all available seasons. More specifically, post-hoc Nemenyi tests indicate that peak timing tended to be later for outbreaks of influenza B ($p < 0.01$ for comparisons against both A(H1) and A(H3)), although we note that the median peak timing for outbreaks of influenza B in our dataset was only 1-2 weeks later than the medians for outbreaks of A(H1) or A(H3).

Because data are not reported as rates, it is difficult to compare peak intensity by subtype. In order to explore whether certain subtypes consistently yielded larger or smaller outbreaks, we first adjusted observed peak intensity values for each country by dividing by the peak intensity of the largest outbreak observed for that country over all seasons and subtypes. We could then assess whether these relative outbreaks sizes differed by subtype. We find that outbreaks of A(H3) influenza tended to be larger than those of A(H1) (post-hoc Nemenyi test, $p < 0.01$) or B

($p < 0.05$). This result is in agreement with previous reports that outbreaks of A(H3) tend to be larger and of higher severity than outbreaks of A(H1) (Park & Ryu, 2018).

Additional Model Fit Results

As described in the main text, model fits for A(H3) and B influenza during the 2012-13 season are visualized in S2 Figure.

Results by Observed Lead Week

Generally, forecasting results were similar whether they were assessed by predicted (Figure 3) or observed (S3 Figure A-C) lead week. This remained true if season-country-subtype-week pairs were removed because either the network or the isolated models did not predict an onset (S3 Figure D-F). Interestingly, if we only include pairs for which both the network and isolated models produce forecasts predicting an onset, the observed improvement in onset timing accuracy forecasts appears to shrink. This suggests that any improvement the network model offers for onset timing predictions is not due to the production of more skillful forecasts, but rather due to the recognition of upcoming outbreak onsets not picked up by the isolated model.

Results by (Sub)type

Results by (sub)type are described in the main text under "Retrospective Forecast Accuracy." Plots of log scores by predicted lead week separated by (sub)type are found in S4 Figure.

Results by Country

Log scores for peak timing and intensity by country by predicted lead week can be seen in S5 Figure. Because so few forecasts of onset timing were produced prior to outbreak onset, we refrain from plotting these results. Although the network model offered no improvement in forecast accuracy overall, we wondered whether improvement may be observed for specific countries. In particular, we might expect that countries with lower-quality data may benefit from the inclusion of other countries with higher-quality data, or that countries with larger commuting flows may benefit more from a network model like the one developed here. However, there is generally very little difference between network and isolated model results in most countries, mirroring the results in the main text. Additionally, Luxembourg, the country with the least-smooth data (see "Data Quality by Country" below), is one of the countries for which the network model clearly degraded peak timing forecast accuracy (the others being Hungary, Italy, and Poland). Czechia and the Netherlands appear to see the most improvement in forecast accuracy when the network model was used, but these improvements are small. Overall, no clear patterns emerged concerning network and isolated model performance by country.

Results (Mean Absolute Errors)

In addition to log scores, we assessed forecast results using mean absolute errors (MAE), a metric which accounts for accuracy but not for certainty or calibration. MAE is calculated by taking the absolute value of the difference between the predicted and observed values of a given metric for all forecasts, then calculating the mean value over all forecasts. Note that, because values for peak intensity varied greatly between countries, we instead calculated mean absolute

percentage errors (MAPE) for peak intensity; here we divide the absolute difference by the observed peak intensity value before taking the mean.

When assessed by predicted lead week, the network model appears to improve forecasts of both peak and onset timing prior to the predicted peak or onset, although we again note that the number of forecasts generated prior to outbreak onsets was small (S6 Figure A and C). Peak intensity, however, appears to be more accurately forecasted by the isolated models, particularly at lead weeks -4 to -2 (S6 Figure B). Thus, results for peak intensity and for onset timing are consistent with those found in the main text using log scores (Figure 3 B and C). Peak timing appears to be more accurately predicted by the network model prior to the predicted peak when assessed using MAE, yet received lower log scores than the isolated model over the same range of lead weeks (Figure 3A). This suggests that network forecasts of peak timing may be closer to the observed values on average, but place less confidence in their forecasts (i.e., fewer ensemble members fall into the observed bin).

MAEs and mean absolute percentage errors (MAPE) separated by (sub)type can be found in S7 Figure. Network model improvements for peak timing forecasts appear greatest for influenza B (S7 Figure C), while improvements for onset timing are primarily found for (sub)type A(H1) (S7 Figure G). As in S6 Figure, the network model appears to degrade peak intensity forecasts, with the potential exception of very early forecasts of A(H3) (S7 Figure E). However, as with the results for all (sub)types combined, the magnitude of these differences is quite small.

Alternative Calibration Plots

If forecasts are properly calibrated, we expect the distribution of forecast errors to have a mean of 0; otherwise, the forecasts are biased. S8 Figure compares the distribution of forecast errors for peak timing and intensity for the network and isolated models. Note again that relative errors were used for peak intensity, to control for differences in intensity by country. We see little difference between the models in terms of bias. Both models appear to produce unbiased estimates of peak timing at all lead week ranges, with the distributions becoming narrower as the peak approaches and passes. Estimates of peak intensity, on the other hand, appear biased low at early lead weeks, but appear relatively unbiased starting at predicted lead week -4, and become more precise at later leads. Overall, these results agree with the finding in the main text that both the network and isolated models are well-calibrated for forecasting peak timing and intensity (Figure 4).

Supplementary Analyses

Synthetic Testing

In order to test whether our model is capable of fitting realistic influenza outbreaks, we conducted synthetic testing, wherein we test the model-inference system's ability to properly fit model-generated "outbreak" data. This is done because, unlike for observed outbreaks, we know the exact parameter and initial states values that were used to generate synthetic data, allowing assessment of the accuracy of the model fitting.

We drew 1000 random combinations of initial state and parameter values from realistic ranges (see main text) using Latin Hypercube Sampling. Unlike for forecasting, initial values of *S* and *I* were chosen for each country rather than for each individual compartment. *S0* in each

commuting compartment was then chosen from a normal distribution around *S0* for those living and working in the commuters' home country with standard deviation 0.025. Initial infected numbers in a country were distributed among all compartments sharing a home country according to relative population size. These state and parameter combinations were then run forward for the duration of an influenza season (52 weeks). Commuting flows were set to be the mean number of commuters across all seasons; missing routes were filled as described in the main text. Realistic outbreaks were determined to be those where: a) at least 7 of 12 countries had outbreak onset (defined as in the main text), and b) no more than 1 of those countries with onsets had peak timings before week 52 or after week 12. These metrics were based on preliminary exploration of the syndromic+ data for each (sub)type. We selected five outbreaks on which to perform synthetic testing. As influenza outbreaks in Europe tend to move from west to east, we chose four outbreaks that progressed from west to east (one strongly and three weakly), and one that moved east to west instead (as observed during the 2015-16 season). Parameter values for these outbreaks can be found in S3 Table.

Because observed data are rarely as smooth as synthetically-generated data, we added random error to our synthetic data. Specifically, we added normally-distributed error to each point with a mean of zero and standard deviation equal to the square root of the observation error variance at that point, calculated as described above under "Ensemble Adjustment Kalman Filter," with b = 1e5 and c = 10. Any resulting negative values were set to zero.

We then fit our network model to each of the 5 error-laden synthetic outbreaks, and compared the inferred and true values of the model parameters D, R_{0max}, and R_{0diff}; and the composite model parameters β, R_0 (as defined in Equation 2 in the main text), and R_{eff} (the effective reproductive number, or the average number of secondary cases caused by a single

initial case, taking population susceptibility into account; calculated as $R_0(\frac{S}{N})$). Because R_{eff} dictates the course of the epidemic, it is particularly important that the model be able to correctly infer its value. Comparisons were achieved by calculating errors relative to the true values at all time points.

Fits for *D, R0max,* and *R0diff* over the course of the five synthetic outbreaks can be found in S9 Figure. We find that the model-inference system fit *D* relatively well in most cases, with a slight tendency to underestimate its value. R_{0max} and R_{0diff} appear more difficult to accurately infer, with the model settling on similar inferred values for all five synthetic outbreaks. This may suggest that the model is not particularly sensitive to values of these two parameters, or that multiple combinations of R_{0max} and R_{0diff} are capable of producing the same outbreak patterns.

The distribution of relative errors for β, R_0, and R_{eff} for all countries and synthetic outbreaks at timepoints 5, 10, 15, and 20 are shown in S10 Figure. Generally, the inferred values of β, R_0, and R_{eff} approached the true values over time. R_0 was fit quite closely at all timepoints (S10 Figure B), while the model seems less capable of correctly inferring β (S10 Figure A). Encouragingly, although R_{eff} tended to be underestimated early in the season, it was well-fit during the outbreak itself (S10 Figure C). That said, the tendency of the model to underestimate R_{eff} before outbreak onset could be why so few early forecasts of onset timing were produced (Table 1).

Data Quality by Country

While small differences in data quality are not necessarily associated with forecast accuracy (Kramer & Shaman, 2019; Morita et al., 2018), we nonetheless expect that particularly poor-quality data will be more difficult to fit, and will produce lower-quality forecasts. We

briefly assessed data quality by country and subtype, and its association with fit and forecast accuracy using two metrics:

1. The proportion of weeks within an outbreak where no data are available (where an outbreak is defined as starting in week 40 and ending in week 19 of the following year)
2. Lag-one autocorrelation by outbreak (a measure of signal smoothness)

For these analyses, only countries with an observed onset were included for each (sub)type and season.

First, we note that the two metrics are not significantly correlated with one another, and in fact trend towards being slightly negatively associated (Kendall's tau = -0.101, $p < 0.07$), emphasizing that there is no one measure that perfectly encapsulates data quality. Missingness (metric 1) ranges from 0 to 37.5% of time points during a season (mean = 9.67%, median = 6.25%), whereas smoothness (metric 2) ranges from 0.29 to 0.95 (mean = 0.83, median = 0.87), suggesting high variability in data quality, and confirming that both missingness and noisiness are prevalent in our data.

Both metrics above differ significantly by country (Kruskal-Wallis test, $p < 0.0001$ for both metrics). Briefly, Germany has particularly low missingness, and Italy has high missingness; Luxembourg has notably low signal smoothness. Because all subtypes make use of the same clinical and virologic data points, it is not meaningful to compare missingness by subtype. Signal smoothness does not differ significantly by subtype (Kruskal-Wallis test, $p = 0.45$).

Finally, we assessed whether either of these metrics were associated with model fit (measured by RMSE) or forecast accuracy (measured by log score). First, RMSE values were

averaged over all runs for each country-season-subtype combination, and log scores were averaged over all runs and lead weeks (predicted leads -6 through 4). Because we are particularly interested in forecast accuracy prior to the peak, we also looked at log scores averaged over predicted leads -6 through -1 only. Neither RMSEs nor log scores were significantly associated with missingness, and only log scores for peak timing were associated with smoothness. Specifically, smoothness was positively associated with peak timing accuracy over the full range of predicted leads (Kendall's tau = 0.112, $p < 0.05$), as well as when only pre-peak leads were included (Kendall's tau = 0.126, $p < 0.02$). We note, however, that these associations were fairly weak. Therefore, at least using these simple metrics of data quality, we find little evidence that data quality differences within our dataset were associated with network model fit quality and forecast accuracy.

Inferred States and Parameters

If a model can only produce appropriate estimates and forecasts by inferring unrealistic values of model states (the number susceptible and infected) and parameters (here, R_{0max}, R_{0diff}, L, D, and *airScale*), it suggests that the model itself may not be well-specified. We therefore assessed whether the values of model states and parameters inferred by the network model were realistic. Additionally, we explored state and parameter patterns by (sub)type. In addition to the model parameters listed, we also looked at inferred values of R_0 (the basic reproductive number), and R_{eff}.

The inferred value of *S0* for a given country during a (sub)type-specific outbreak was taken to be the maximum inferred proportion susceptible over the time period beginning with the first onset observed in that (sub)type-specific outbreak, and ending in week 19 (i.e., the end of

the season). Weeks prior to the first onset were removed because, in synthetic testing (see above), the proportion susceptible was often overestimated during the first few weeks of fitting, particularly in countries with low *S0*. The maximum R_{eff} for a country during a given (sub)type-specific outbreak was found using the same method. R_0 was taken as the value of R_0 during the week with maximum R_{eff}. Because our model assumes a single pathogen, we expect R_0 to vary between countries only according to differences in absolute humidity. Thus, after confirming that no significant differences existed for R_0 by country (Kruskal-Wallis test, as described in next paragraph; all $p > 0.05$), we calculated an overall estimate of R_0 for each (sub)type-specific season by taking the mean R_0 over all countries. The five model parameters were simply assessed at week 7, twenty weeks after fitting begins.

Each season and (sub)type was fit five times, with each run having different initial conditions and commuting matrices (see main text). Because these runs were not independent, we randomly chose a single run for each country-season-(sub)type combination (season-(sub)type combinations for R_0 and model parameters) before checking for significant differences using Kruskal-Wallis rank sum tests. This process was repeated 100 times, and differences were considered significant if at least 50 of the random permutations yielded p-values less than 0.05 (see S1 Text from (Kramer & Shaman, 2019)).

Estimates for *S0* ranged from 54.7% to 95.2% of a country's population (58.0% to 95.2% among those countries, seasons, and (sub)types where an onset was observed), and maximum R_{eff} fits ranged from 0.68 to 2.10 (0.93 to 2.10 where onsets occurred). We found that estimates of maximum R_{eff} differed significantly by (sub)type (S11 Figure B). Post-hoc Nemenyi tests using a Bonferroni correction for multiple comparisons, also performed on 100 permutations of the estimates, revealed that inferred maximum R_{eff} values were significantly higher for A(H3) than

for A(H1), results that are in line with our finding above that A(H3) outbreaks in our dataset tended to be larger than those of other (sub)types (see "Descriptive Statistics by (Sub)type"). While the initial Kruskal-Wallis test suggested a significant difference in *S0* by country, post-hoc Nemenyi tests found that none of the pairwise comparisons were significant once a Bonferroni correction was applied. That said, Kendall's rank correlation suggested that countries that are further east and further north (using the latitude and longitude of each country's capital city) tended to have lower *S0*; this pattern may reflect the model's attempt to fit outbreaks that tend to move from west to east. No significant differences in *S0* were observed by (sub)type (S11 Figure A).

Generally, parameters common to all countries also fell within realistic ranges (Carrat & Flahault, 2007; Chowell et al., 2008; Mills et al., 2004; Truscott et al., 2009; White & Pagano, 2008), with values ranging from 1.23 to 2.37 for R_0, from 2.56 to 10.93 days for *D*, and from 4.8 to 9.5 years for *L*. Estimated values for *airScale* fell between 0.82 and 1.33, although we note that neither *airScale* nor *L* were inferred particularly well in synthetic testing, and results for these parameters should be interpreted with caution. Using Kruskal-Wallis tests as described above, none of these parameters were significantly associated with (sub)type. However, we note that, since only a single value of each parameter is estimated for each season-(sub)type combination, the "sample size" here is quite small. In particular, there seems to be a trend toward higher estimates for *D* during outbreaks of A(H3) and B as compared to A(H1) (S11 Figure D).

References

Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. *Vaccine, 25*(39–40), 6852–6862. https://doi.org/10.1016/j.vaccine.2007.07.027

Chowell, G., Miller, M. A., & Viboud, C. (2008). Seasonal influenza in the United States, France, and Australia: Transmission and prospects for control. *Epidemiology and Infection*, *136*(06). https://doi.org/10.1017/S0950268807009144

Decoville, A., Durand, F., Sohn, C., & Walther, O. (2013). Comparing Cross-border Metropolitan Integration in Europe: Towards a Functional Typology. *Journal of Borderlands Studies*, *28*(2), 221–237. https://doi.org/10.1080/08865655.2013.854654

EU labour force survey—Main features and legal basis. (n.d.). Eurostat. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_force_survey_%E2%80%93_main_features_and_legal_basis

European Commission. (2003). *The European Union labour force survey—Methods and definitions—2001*. European Communities. https://unstats.un.org/unsd/EconStatKB/Attachment269.aspx?AttachmentType=1

Kramer, S. C., & Shaman, J. (2019). Development and validation of influenza forecasting for 64 temperate and tropical countries. *PLOS Computational Biology*, *15*(2), e1006742. https://doi.org/10.1371/journal.pcbi.1006742

Mathä, T., & Wintr, L. (2009). Commuting flows across bordering regions: A note. *Applied Economics Letters*, *16*(7), 735–738. https://doi.org/10.1080/13504850701221857

Mills, C. E., Robins, J. M., & Lipsitch, M. (2004). Transmissibility of 1918 pandemic influenza. *Nature*, *432*(7019), 904–906. https://doi.org/10.1038/nature03063

Morita, H., Kramer, S., Heaney, A., Gil, H., & Shaman, J. (2018). Influenza forecast optimization when using different surveillance data types and geographic scale. *Influenza and Other Respiratory Viruses*. https://doi.org/10.1111/irv.12594

Park, J.-E., & Ryu, Y. (2018). Transmissibility and severity of influenza virus by subtype. *Infection, Genetics and Evolution*, *65*, 288–292. https://doi.org/10.1016/j.meegid.2018.08.007

Pei, S., Kandula, S., Yang, W., & Shaman, J. (2018). Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences*, 201708856. https://doi.org/10.1073/pnas.1708856115

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3837

Truscott, J., Fraser, C., Hinsley, W., Cauchemez, S., Donnelly, C., Ghani, A., Ferguson, N., & Meeyai, A. (2009). Quantifying the transmissibility of human influenza and its seasonal variation in temperate regions. *PLoS Currents*, *1*, RRN1125. https://doi.org/10.1371/currents.RRN1125

White, L. F., & Pagano, M. (2008). Transmissibility of the influenza virus in the 1918 pandemic. *PloS One*, *3*(1), e1498. https://doi.org/10.1371/journal.pone.0001498

Supplementary Tables

S1 Table. Scaling factors by country and (sub)type. Note that France switched from preferentially collecting ARI to ILI data prior to the 2014-15 season; scaling factors for France are listed as ARI/ILI.

Country	Scaling (A(H1))	Scaling (A(H3))	Scaling (B)
Austria	17.53	25.09	23.30
Belgium	12.98	15.81	15.92
Czechia	1.22	1.32	1.33
France	2.08/0.074	3.92/0.033	1.22/0.12
Germany	0.32	0.41	0.50
Hungary	1.81	1.18	1.19
Italy	2.04	1.75	1.19
Luxembourg	83.24	74.33	82.04
Netherlands	32.32	46.89	42.13
Poland	2.56	1.68	1.65
Slovakia	0.58	1.17	0.81
Spain	3.27	5.16	6.86

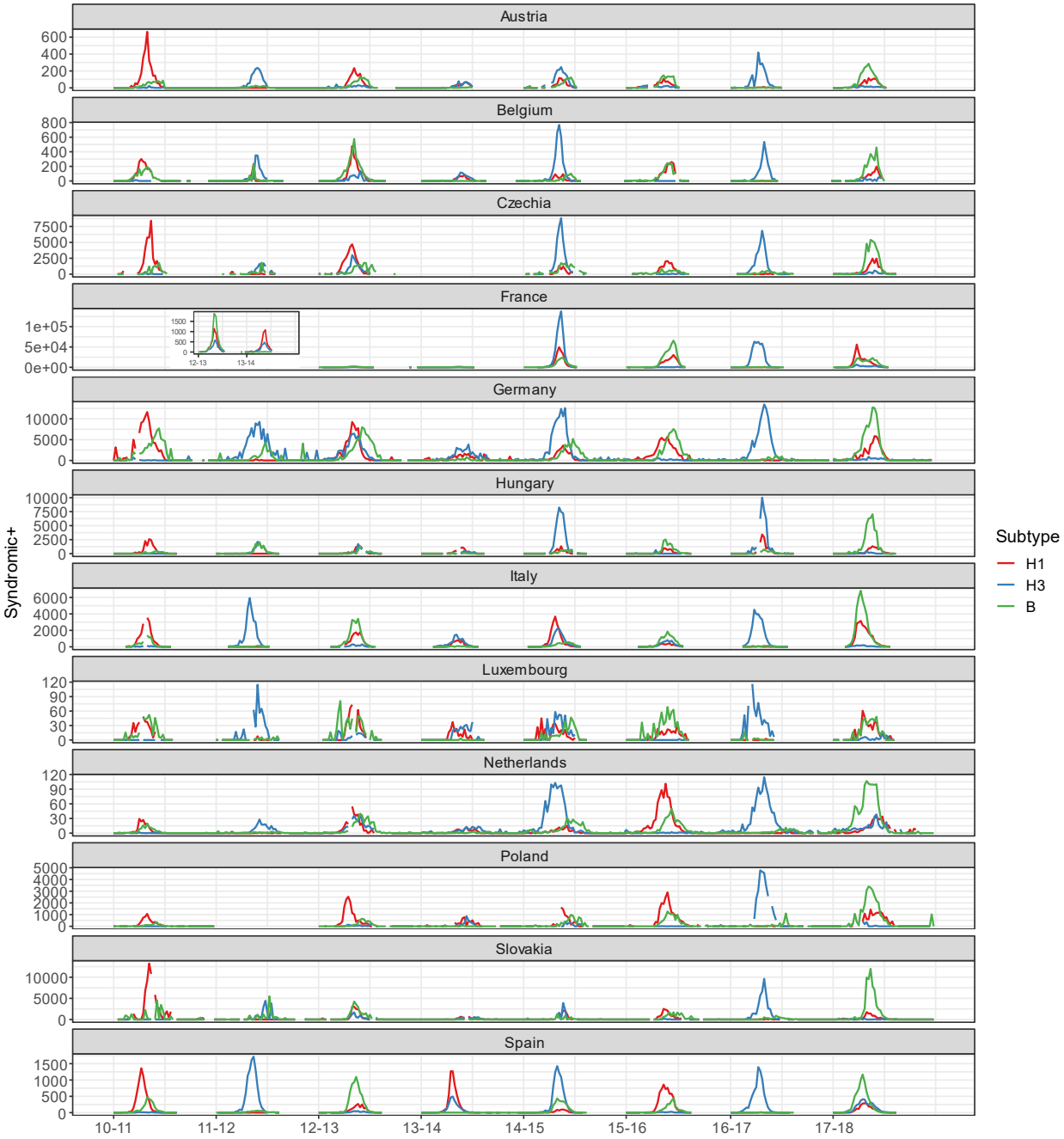
S2 Table. Number of forecasts predicting any onset by predicted onset lead week, separated by (sub)type.

	Lead Week	-6	-5	-4	-3	-2	-1	0
A(H1)	Network	1	1	2	12	15	26	239
	Isolated	6	22	11	25	16	10	205
A(H3)	Network	1	14	9	10	30	40	230
	Isolated	3	11	4	13	18	10	186
B	Network	0	2	1	9	18	50	227
	Isolated	2	10	5	10	10	17	229

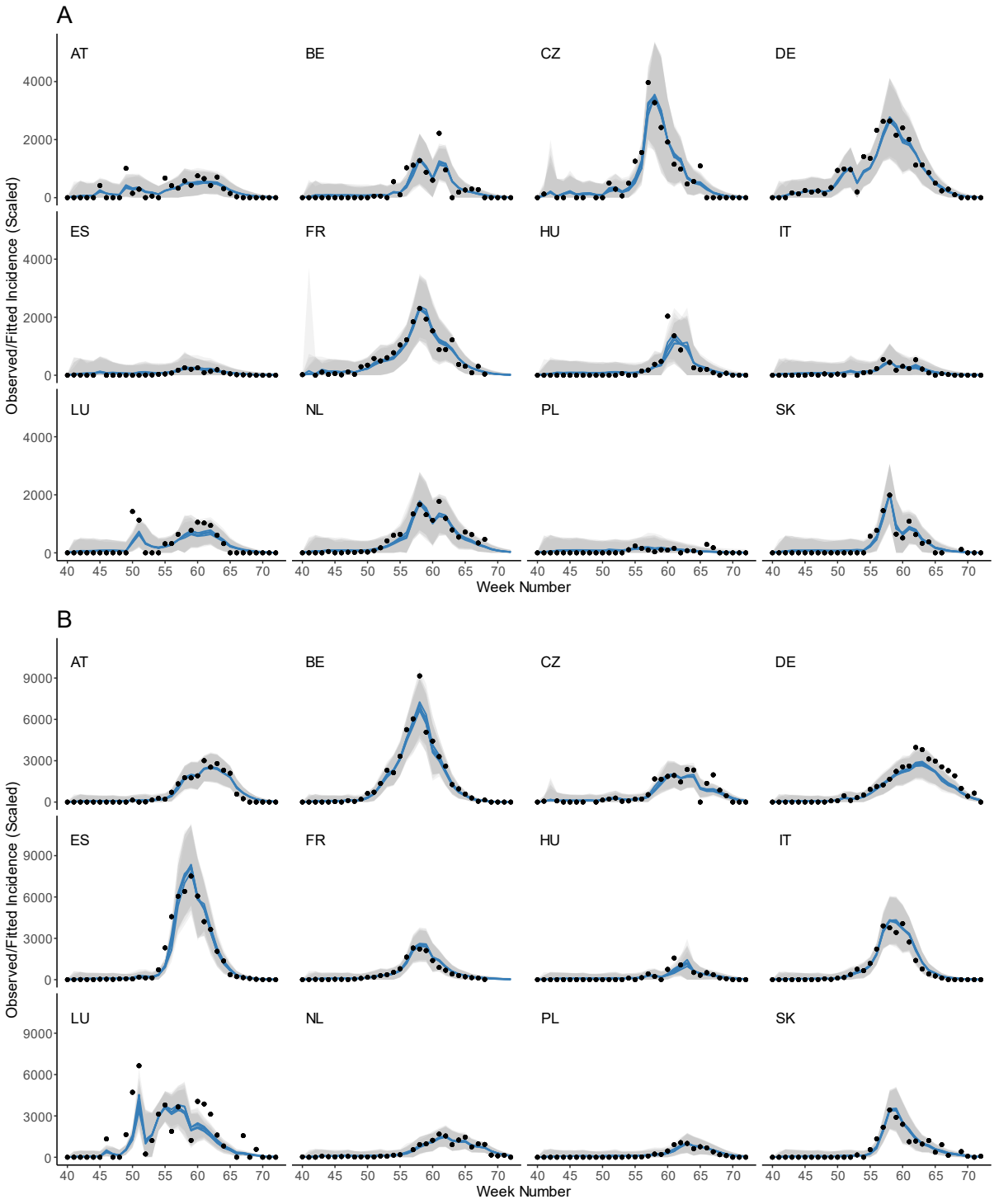
S3 Table. Parameters values used to generate synthetic outbreaks, and the number of countries with outbreak onsets.

Synthetic Outbreak	*L* (years)	*D* (days)	R_{0max}	R_{0diff}	*airScale*	# of Countries w/ Onsets
1	4.61	5.58	2.15	0.48	1.24	10
2	3.36	6.18	2.28	0.57	0.92	11
3	4.92	6.14	2.63	0.98	1.17	12
4	8.56	3.96	2.11	0.90	0.87	7
5	6.33	4.71	2.33	0.84	1.16	11

Supplementary Figures

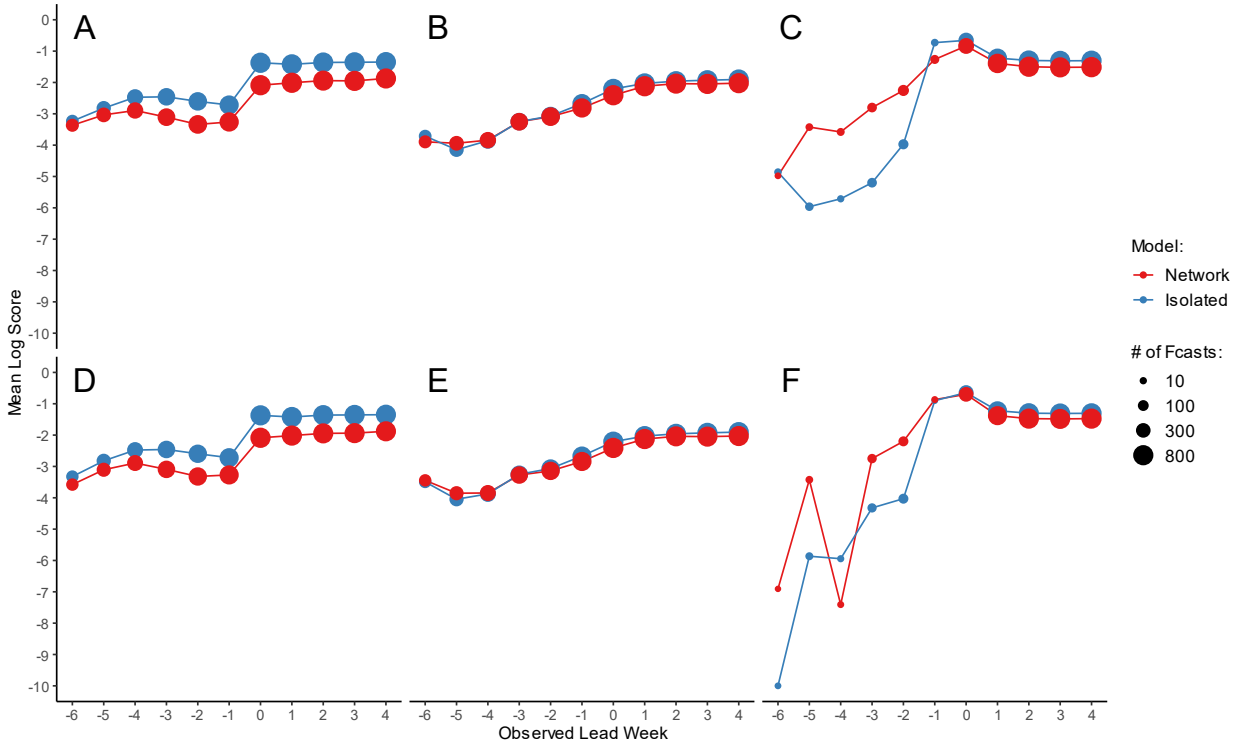


S1 Figure. Syndromic+ data by country and subtype over the course of the study period. The beginning of each season (week 40) is marked on the x-axis. Because France shifted from preferentially reporting ARI to ILI data prior to the 2014-15 season, ARI data from the 2012-13 and 2013-14 seasons are shown as an inset.



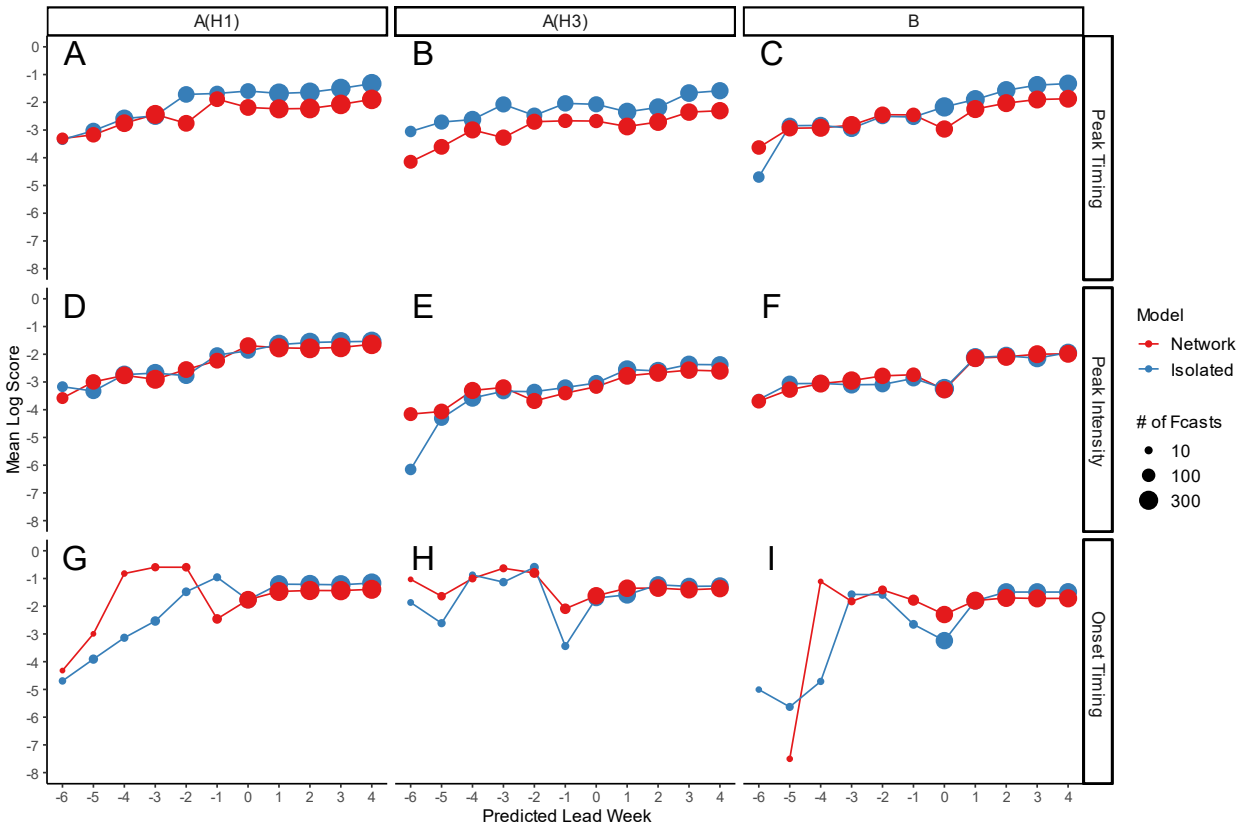
S2 Figure. Model fitting to observed A(H3) (A) and B (B) influenza data throughout the 2012-13 season.

Scaled syndromic+ data are shown as points, while blue lines represent inferred model incidence. Each line represents one of five runs, each with different initial conditions. The shaded gray areas represent the 95% confidence intervals for each of the runs, calculated using the ensemble means and standard deviations. (A) RMSE ranged from 45.60 (ES) to 375.08 (CZ), with mean = 180.45 (median = 166.77). (B) RMSE ranged from 106.01 (PL) to 1155.82 (LU), with mean = 364.01 (median = 309.06).

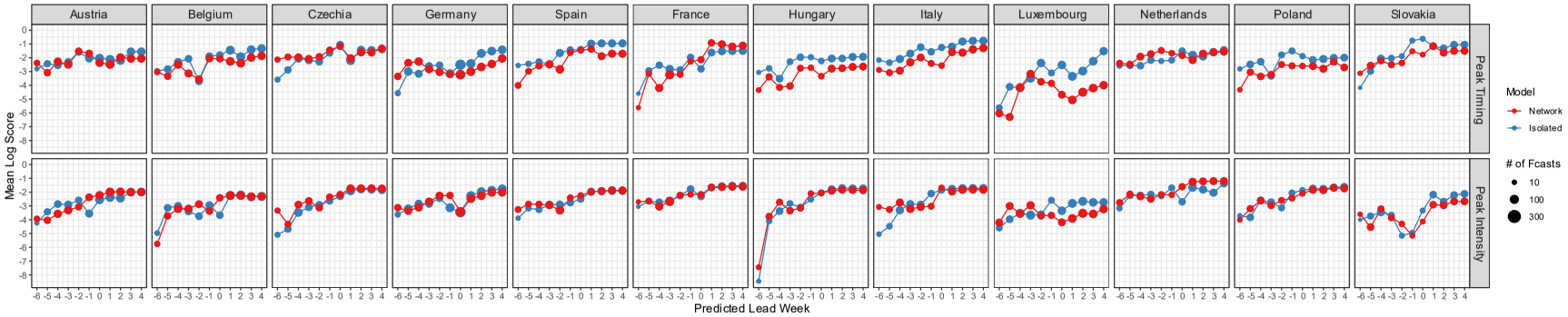


S3 Figure. Retrospective forecast accuracy by observed lead week, before (A-C) and after (D-F) removing season-country-subtype-week pairs for which either the network or isolated model predicted no onset.

Log scores are shown for peak timing (A and D), peak intensity (B and E), and onset timing (C and F). Mean log scores for the network model are shown in red, and scores for the isolated models are shown in blue. The size of the points represents the number of forecasts generated at a given lead week for which any onset was predicted.

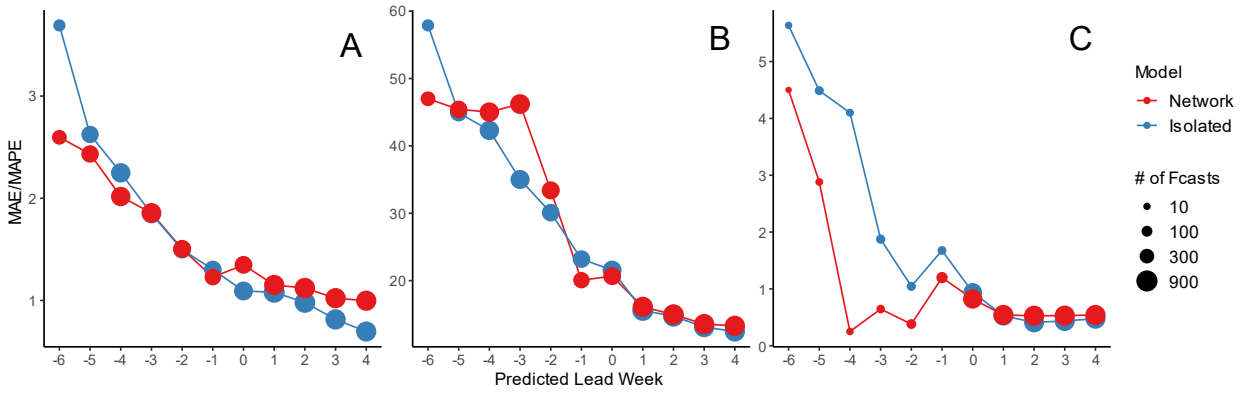


S4 Figure. Log scores by predicted lead week for forecasts of peak timing (A-C), peak intensity (D-F), and onset timing (G-I), separated by (sub)type.
Network results are shown in red, and results from the isolated model are shown in blue. Point size represents the number of forecasts generated at a given lead week for which an onset was predicted.

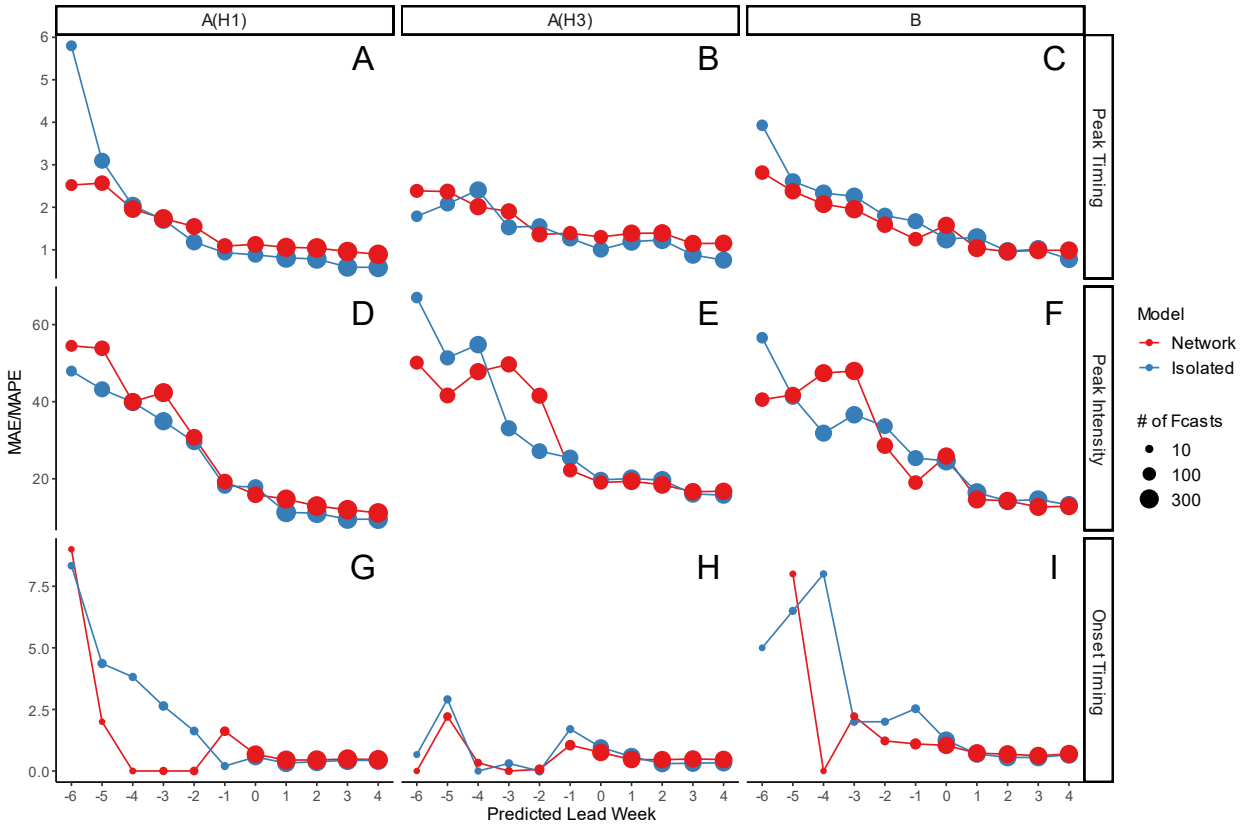


S5 Figure. Log scores by predicted lead week for forecasts of peak timing and peak intensity, shown separately for each of the twelve countries in the network model.

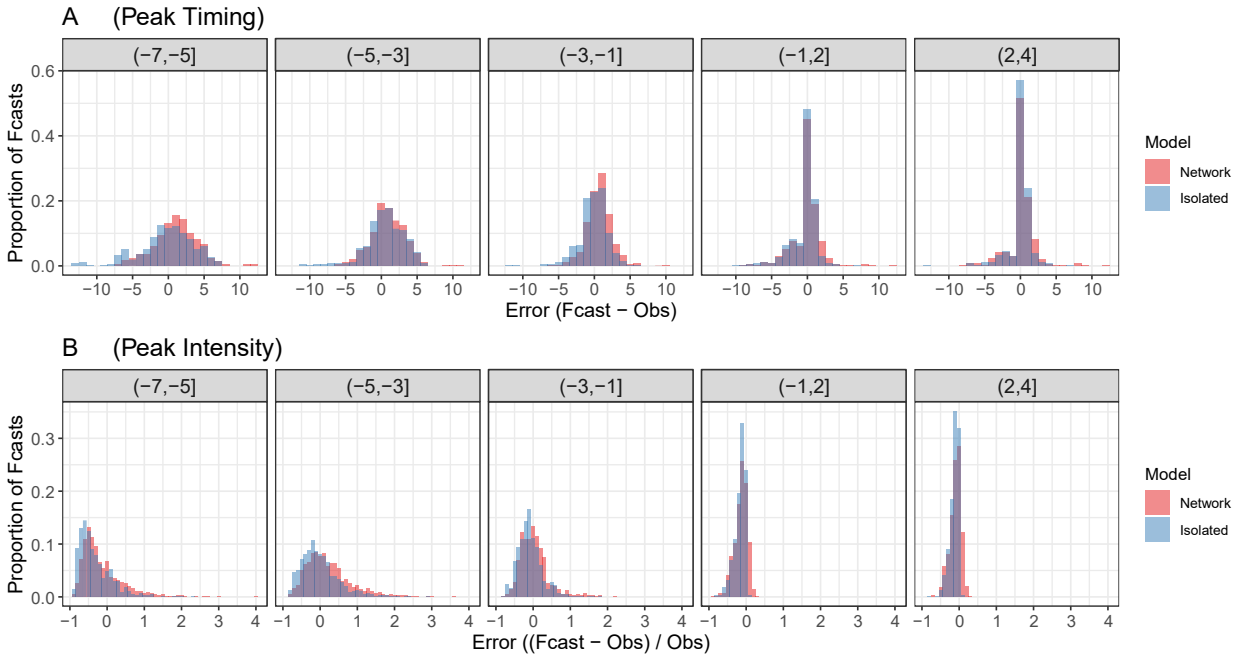
Network results are shown in red, and isolated in blue. Point size represents the number of forecasts generated that predicted an onset.



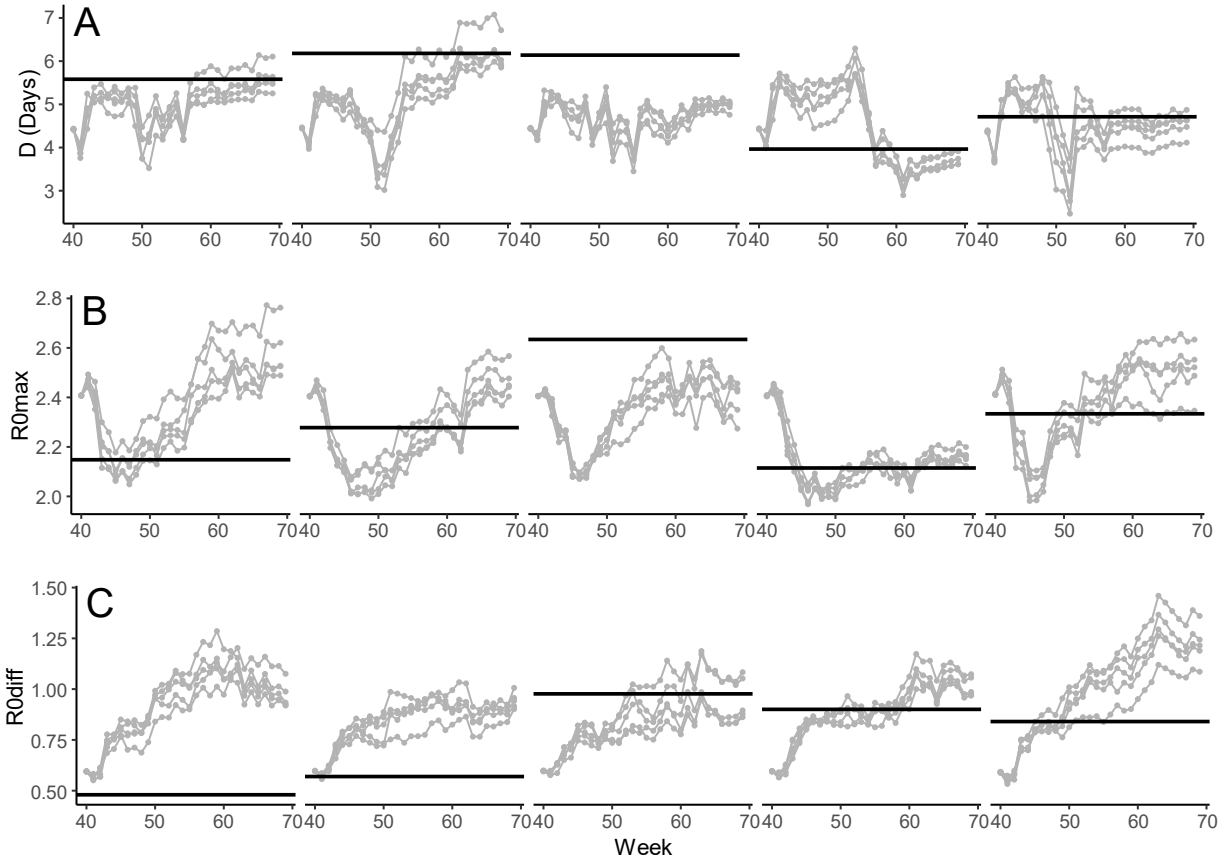
S6 Figure. Retrospective forecast accuracy by mean absolute (percentage) errors (MA(P)E).
Results are shown by predicted lead week for peak timing (A), peak intensity (B; MAPE), and onset timing (C). Network results are in red and isolated in blue. Point size represents the number of forecasts generated for which onsets were predicted.



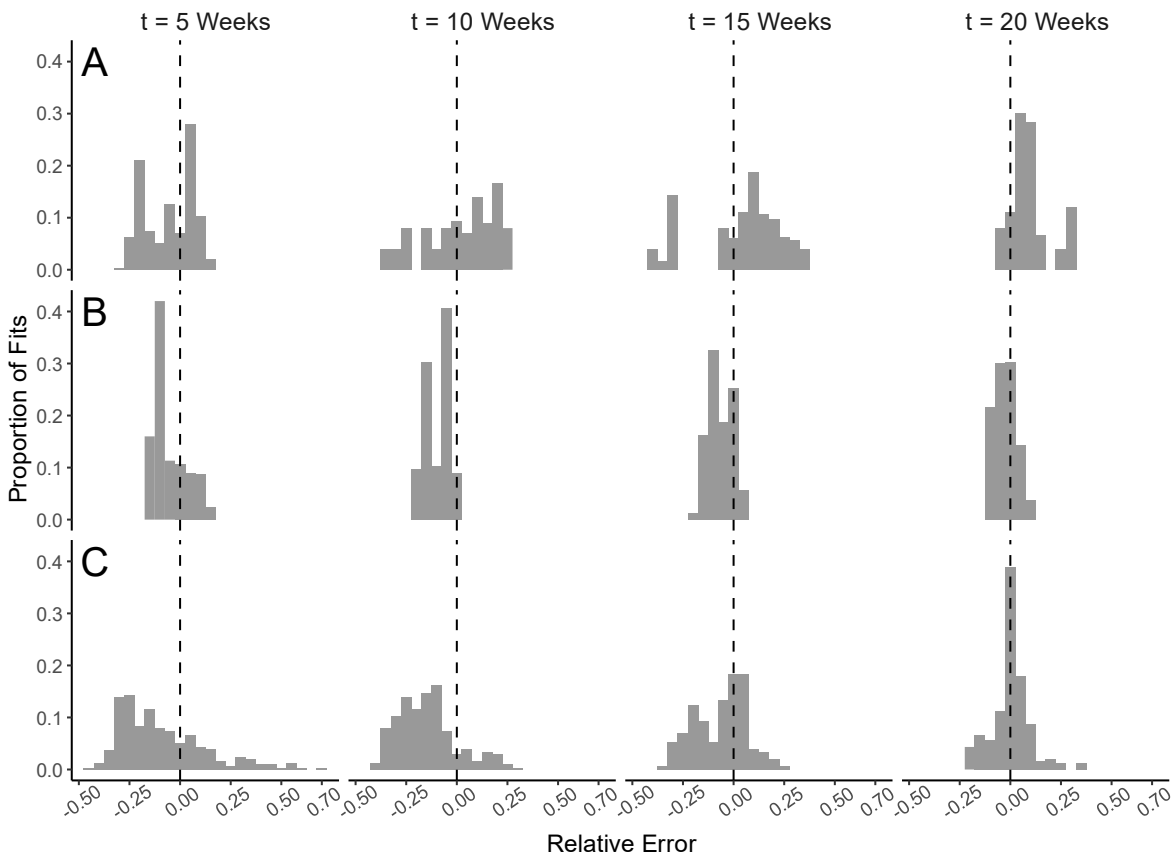
S7 Figure. Retrospective forecast accuracy by MAE/MAPE by predicted lead week, separated by influenza (sub)type.
Results are shown for peak timing (A-C), peak intensity (D-F), and onset timing (G-I). Network results are in red and isolated in blue. Point size represents the number of forecasts generated for which an onset was predicted.



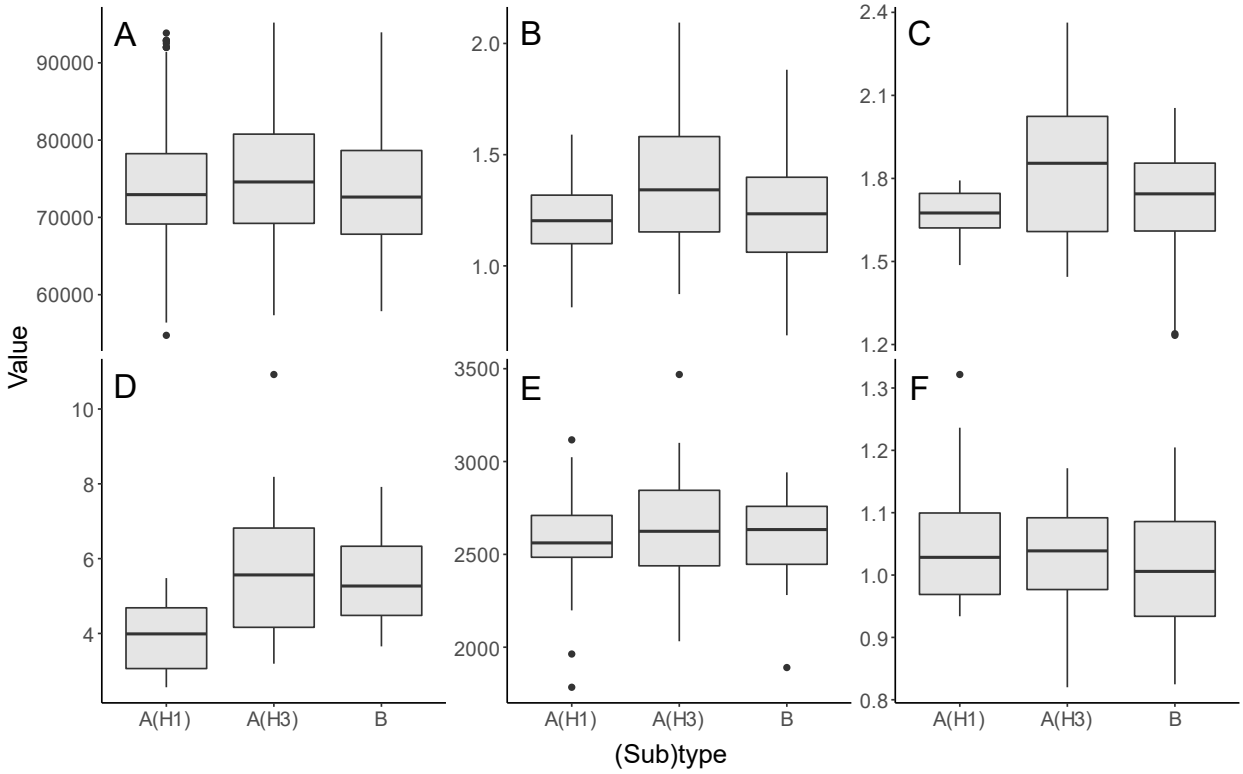
S8 Figure. Histograms of forecast error for peak timing (A) and intensity (B), shown by (binned) predicted lead week.
Peak timing error is shown with bins of size 1 week, and relative peak intensity error is shown with bins of size 0.1. The y-axis represents the proportion of forecasts falling into a given bin. Results from the network model are shown in red; results from the isolated model are shown in blue.



S9 Figure. Fittings of model parameters *D* (A), *R0max* (B), and *R0diff* (C) over time.
Each panel in (A), (B), and (C) represents one of each of five synthetic outbreaks, as described in S3 Table. Fits at each week are shown as gray points; solid black lines represent the true values.



S10 Figure. Histograms of the relative error of fits of β (A), R_0 (B), and R_{eff} (C) at four timepoints over all countries and synthetic outbreaks.
The y-axis shows the proportion of fits falling into each bin of size 0.05. Dotted lines show relative error equal to 0, for reference.



S11 Figure. Values of model states and parameters, as fit by the network model, shown separated by (sub)type.
Fittings are shown for *S0* (A), maximum *Reff* (B), *R0* (C), *D* (D), *L* (E), and *airScale* (F). Boxes extend from the first to the third quartile, with a horizontal line marking the median, while values more than 1.5 times the interquartile range below the first quartile or above the third quartile are shown as points.

Chapter 4

Real-time Forecasting of the 2017-18 and 2018-19 Influenza Seasons in 37 Countries

Acknowledgments:
I would like to thank Jeffrey Turmelle for creating and maintaining the real-time forecasting website, and for his patience and assistance as we worked to get the new forecasts online. I would also like to thank Sasikiran Kandula for helping with much of the code needed to format and submit the forecasts for online viewing. Thank you also to Ralf Strobel for assisting with figure formatting. Finally, I would like to thank participants of the Influenza Incidence Analytics Group calls for helpful comments and suggestions.

Abstract

Accurate and well-calibrated influenza forecasts, generated in real-time, have the potential to guide public health preparations for unfolding influenza outbreaks, thereby reducing morbidity and mortality. While these systems are becoming more established in the United States and Australia, real-time forecasting systems have not been developed for most countries worldwide. Here we describe the generation of real-time forecasts for 37 countries over the course of two influenza seasons, which were published weekly online. Here we show that, while forecast quality varied by season and country, forecast skill was on par with that of previously published retrospective forecasts. Forecasts of peak intensity in particular were found to outperform methods relying on historical expectance alone in the weeks directly before peak influenza incidence. Furthermore, updates made to the data over the course of the season were not found to negatively influence forecast accuracy. While evidence from future seasons will be critical in confirming these findings, our results suggest that real-time influenza forecasts could begin to be operationalized for several countries in Europe.

Introduction

Influenza causes regular, seasonal outbreaks each year in temperate regions (Bloom-Feshbach et al., 2013). Although we can confidently say that these outbreaks will occur during the winter, the exact progression of a given outbreak is not known. Accurate, real-time forecasts of influenza outbreaks, if produced with sufficient lead time, could allow both public health and medical professionals to better prepare for specific outbreaks of influenza as they occur.

Previous work has shown that skillful forecasts of influenza activity are possible (Hickmann et al., 2015; Moss et al., 2017; Nsoesie et al., 2014; Pei et al., 2018; Shaman & Karspeck, 2012; Yang et al., 2015). However, with a few exceptions (Farrow et al., 2017; Moss et al., 2018; Ong et al., 2010; Shaman et al., 2013), most of these studies have been performed retrospectively; in other words, forecasts have been generated for influenza outbreaks that have already occurred. Forecasting in real-time presents additional challenges, in that it relies on prompt and accurate data reporting. Since the 2013-14 influenza season, the Center for Disease Control and Prevention (CDC) in the United States has hosted an influenza forecasting challenge, in which various research groups submit real-time influenza forecasts, which are then scored and compared (Biggerstaff et al., 2016). In addition to participating in these challenges, our group has been producing real-time forecasts for all 50 US states, as well as around 80 US cities, since 2012, and has published these results online since 2013 (Columbia University Mailman School of Public Health, n.d.). In Melbourne, Australia, Moss et al. (2018) have been collaborating with the local Health Department since 2015 to improve real-time influenza forecasts based on discussions between modelers and those in charge of surveillance.

We have previously shown that skillful retrospective forecasts can be generated for several countries in temperate regions using publicly-available data from the World Health

Organization (WHO) (Kramer & Shaman, 2019, or Chapter 2). Here, we expand these efforts to produce real-time forecasts in 37 temperate countries. We explore the timeliness of reporting to the WHO, and compare real-time and retrospective forecasts to quantify the extent to which data updates over time impact forecast accuracy. We also compare our real-time forecasts to forecasts based on historical expectance. Given previous findings, we hypothesize that our forecasts will provide additional information over historical expectancy alone, but that this will depend on the timely and accurate submission of surveillance data.

Methods

Influenza Data

Country-level influenza data were downloaded weekly from the WHO's FluNet and FluID platforms, which collect virologic and epidemiologic surveillance data, respectively (WHO, n.d.-b, n.d.-a). Epidemiologic surveillance data consisted of cases of influenza-like illness (ILI) in 24 countries and acute respiratory infection (ARI) in 13 countries. A map of countries with available data and the type of epidemiologic data reported can be seen in Figure 1. As in past work, we multiply ILI and ARI cases by the proportion of tests conducted during a given week that were positive for influenza in order to control for lack of specificity (Kramer & Shaman, 2019; Shaman et al., 2013). We refer to the resulting data as ILI+ or ARI+. We note that data are reported to the WHO with a one-week lag, meaning that data downloaded each week reflect influenza activity during the previous week.

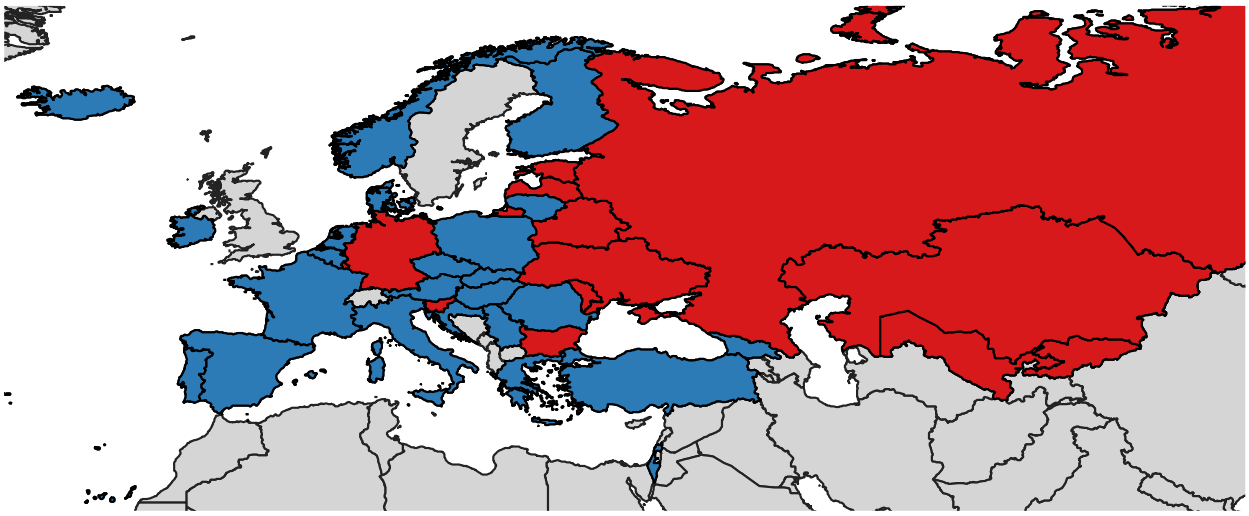


Figure 1. Countries submitting data in real-time, by syndromic data type. This map shows the 37 countries for which real-time forecasting was attempted for at least one season. Countries shown in blue report ILI data, while countries in red report ARI data.

Humidity Data

Data on absolute humidity were obtained from NASA's Global Land Data Assimilation System (GLDAS) (Rodell, n.d.). The data were processed and aggregated to the country level as described in Chapter 2.

Real-Time Forecast Generation

Real-time forecasts were generated using the model-inference system described in Chapter 2. Briefly, a humidity-forced, compartmental SIRS model was fit to available data for each country using the Ensemble Adjustment Kalman Filter (EAKF), a Bayesian data assimilation method (Anderson, 2001). Then, using the estimated model states (number of people susceptible and infected) and parameters (transmission rate, recovery rate, rate of immunity loss), the model was run forward in free simulation through the end of the season. As described previously, we repeat this process five times for each country, each time with different initial conditions, in order to account for stochasticity. Each forecast is the mean of 300

individual ensemble members, which allows assessment of the certainty and calibration of the forecasts (Pei et al., 2018; Shaman et al., 2013; Shaman & Karspeck, 2012; Yang et al., 2015). This process was conducted in real-time, such that the forecasts for each week reflected the state of data availability that week. For the sake of consistency, we began generating forecasts in week 48 (early October) and stopped at week 17 of the following year (mid April) for both seasons. Forecasts were not produced if a country reported 0 cases for a given week, as counts of zero early in an outbreak are unlikely to yield accurate forecasts (historical expectance is more accurate at that stage), and zeros during an outbreak were deemed to be unreliable reporting errors.

Retrospective Forecasts

Even when data, e.g. ILI+, are made available in a timely manner, these data are not always final and may be updated in subsequent weeks over the course of an outbreak. It is consequently not unusual for retrospective forecasts, which use more accurate, finalized data, to outperform real-time forecasts. To test this effect of data updating, we generated retrospective forecasts for each season, using data downloaded at the end of the full 52-week season in question (i.e., week 39 of 2018 for the 2017-18 season and of 2019 for the 2018-19 season). These forecasts were generated using the same methods as described above and in Chapter 2.

Scaling Factors

While model output represents the number of new influenza cases in a given week, our data estimate the number of people with influenza who sought medical treatment. In order for our model-inference system to function, data and model output must be in the same form.

Additionally, because we are using count, not rate, data, we must also account for differences in population size and the size of surveillance catchment areas. This was accomplished using country-level scaling factors, calculated as described in Chapter 2. Because these scaling factors are determined based on the range of attack rates observed in previous seasons, scaling factors for real-time forecasts are updated at the end of each new season, as needed. Country-level scaling factors for each season can be found in S1 Table.

Forecast Evaluation

Forecasts were evaluated based on their ability to accurately predict peak timing (i.e., the week at which influenza incidence was highest) and peak intensity (i.e., influenza incidence at the peak week). Specifically, forecasts were considered accurate for peak timing if the mean peak week predicted by the 300 ensemble members fell within one week of the observed week, and forecasts were considered accurate for peak intensity if the mean predicted intensity fell within 25% of the observed value. Observed peak timing and intensity values were calculated based on the data downloaded at the end of the season. These metrics and thresholds are the same as those used in our recent work generating retrospective forecasts for these countries (see Chapter 2), allowing for comparison of the results.

Forecast Comparison

As in Chapter 2, forecast accuracy by season, data type, and region (S1 Table lists countries by data type and region) was compared using generalized estimating equations (GEEs), which are able to control for temporal autocorrelation between forecasts for the same country-season pair over time. Because the observed time to the peak cannot be known in real-time, we

compared forecasts by predicted lead week, or the difference between the current week and the predicted peak timing. Forecasts for which no onset was predicted were removed from consideration, where outbreak onset was defined as the first of three consecutive weeks with scaled incidence over 500. To control for the lack of independence between the five runs of each forecast, we ran GEEs on 100 permutations of our results, each time randomly choosing a single run of each forecast. The median coefficients and standard errors were then used to calculate the adjusted odds ratios and 95% confidence intervals (see the Supplementary Materials from Chapter 2 for more details).

Because real-time and retrospective forecasts were generated for the same countries, seasons, and forecast start weeks, we chose to analyze these results by observed lead week using exact binomial tests for lead weeks -6 through 4, with a Bonferroni correction to account for multiple tests. Here, forecasts incorrectly predicting no onset were treated as "inaccurate," in order to avoid discounting forecast pairs where one accurately anticipated epidemic activity and the other did not.

Processing for Display on Website

Ideally, we would only display forecasts believed to be accurate and certain enough to be informative. However, waiting too long to display results prohibits early forecasts from offering what information they can. To balance these needs, we began displaying forecasts online as soon as scaled incidence for a country exceeds 300 cases. However, we also displayed the distribution of predicted onset timing, peak timing, and peak intensity over all 300 ensemble members, to enable users to assess forecast certainty. Finally, to enable comparison between countries, we reported intensity relative to the historical maximum incidence observed for a country since the

beginning of the 2010-11 season. Like scaling factors, the historical maximum values are updated yearly as needed.

Results

Description of website

Screenshots of the real-time forecasting website (Columbia University Mailman School of Public Health, n.d.) can be found in Figure 2. Current influenza incidence (here at week 60, or mid-February, of the 2019-20 season), as a percentage of the historical maximum incidence, can be observed under the "World Map" tab (Figure 2A). Hovering over a country provides the user with the exact percentage of the historical maximum, along with the unscaled ILI+ or ARI+ incidence for that country. Maps for previous weeks within the same season can be observed using the "Select Week" drop-down menu.

Clicking on a country leads to a page resembling Figure 2B (here shown for Poland, also at week 60). The top plot shows the observed, unscaled incidence so far, with the predicted peak labeled with a vertical red line. The mean forecast trajectory for that week is shown plus or minus the standard deviation across all ensemble members (the shaded region). Exact values can be seen by hovering over the points. Forecasts generated at previous weeks can be observed simultaneously using the "History Off/On" switch. Predicted peak timing, peak intensity, and onset timing, along with standard deviation, are found in the upper left corner. The middle set of plots displays the distribution of ensemble members predicting various values for onset timing, peak timing, and peak intensity. Hovering over a bar provides the estimated probability associated with that week or intensity range. Finally, the bottom plot displays unscaled influenza incidence for the country over all previous seasons with available data.

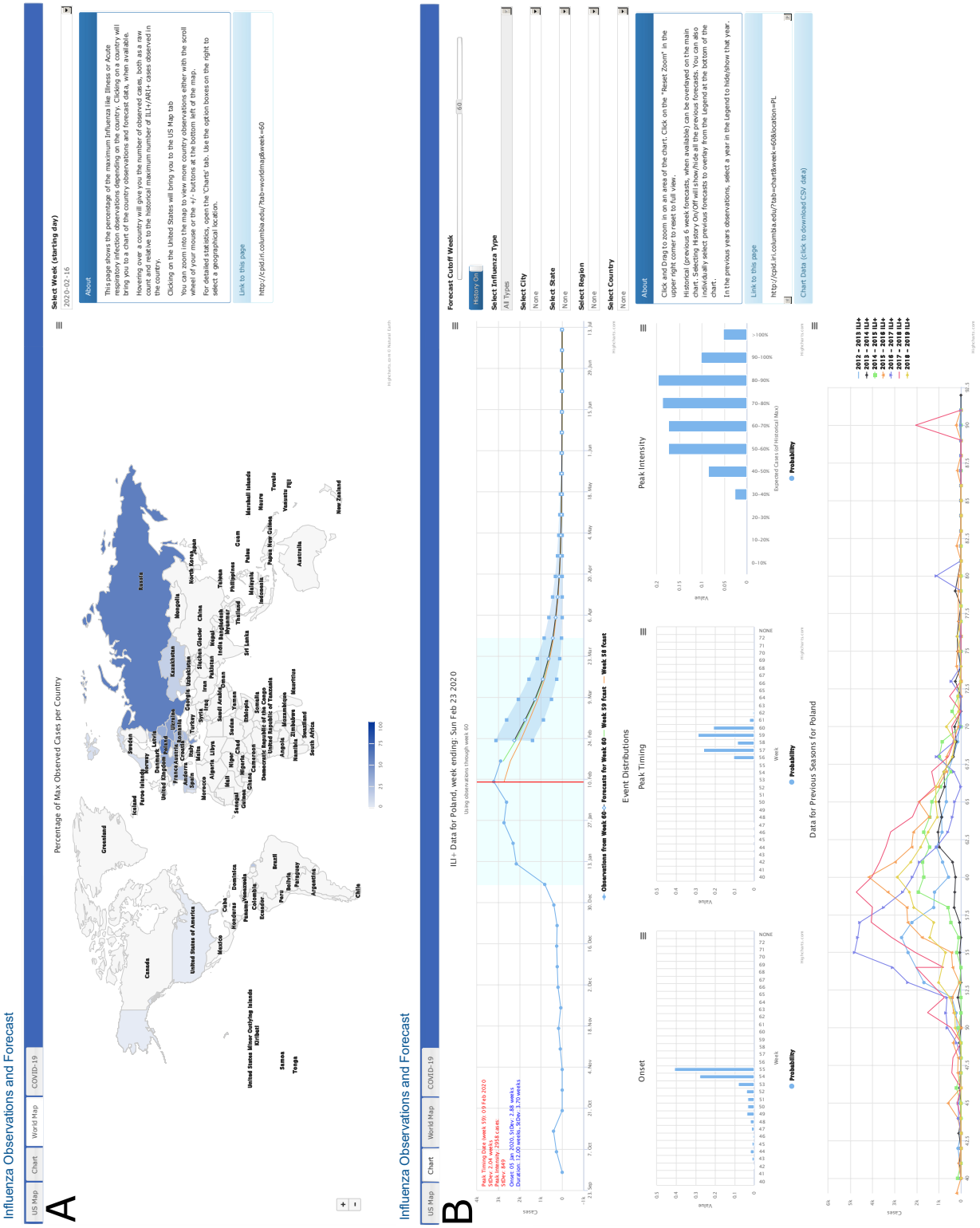


Figure 2. The real-time influenza forecasting website interface. (A) World map showing country-level ILI and ARI incidence, relative to each country's historical maximum, for week 60 of the 2019-20 season. (B) Real-time forecasting results for Poland for the same week, including the mean trajectory, distribution of timing and intensity metrics, and historical incidence. Additional details can be found under "Description of website."

Description of seasons

Reliable and consistent data were available for 37 countries during the 2017-18 season and 35 countries (all except the Netherlands and Uzbekistan, which reported only early in the season or only sporadically, respectively) during the 2018-19 season. Georgia did not have an outbreak onset during the 2017-18 season, while all countries with available data had outbreaks during the 2018-19 season.

Both seasons in Europe were characterized by moderate but varied intensity. The 2017-18 season had somewhat early onset but longer duration of activity than in previous seasons. The timing of the 2018-19 season was in line with past seasons but varied noticeably by country. Interestingly, the 2017-18 season in Europe was dominated by influenza B; during the 2018-19 season, both H1N1 and H3N2 circulated, with H1N1 dominating in the east and north, and both subtypes circulating in the west (Hammond et al., 2018, 2019).

Timeliness of data reporting

During the 2017-18 and 2018-19 influenza seasons, the percentage of weeks during which a country reported no data, making real-time forecasting impossible, ranged from 0 to 0.909 (mean = 0.251; median = 0.205). If reports of 0 cases are included, these numbers rise to 0 to 0.955 (mean = 0.312; median = 0.273). Consistently high rates of reporting were observed among Luxembourg, Portugal, and Slovakia, while the highest rates of missingness were observed for Kyrgyzstan (>80% for both seasons). Only Portugal and Spain achieved reporting every week for the 2018-19 season. During the 2017-18 season, there were two weeks (52 and 64) where no country reported data. We found no significant differences in data missingness by season, both excluding and including reports of 0 cases (Kruskal-Wallis test, $p > 0.35$ for both

tests). Rates of missingness were much lower in the data used for retrospective forecasting, although some missingness persists (mean = 0.024 excluding 0s, mean = 0.101 when 0s are included).

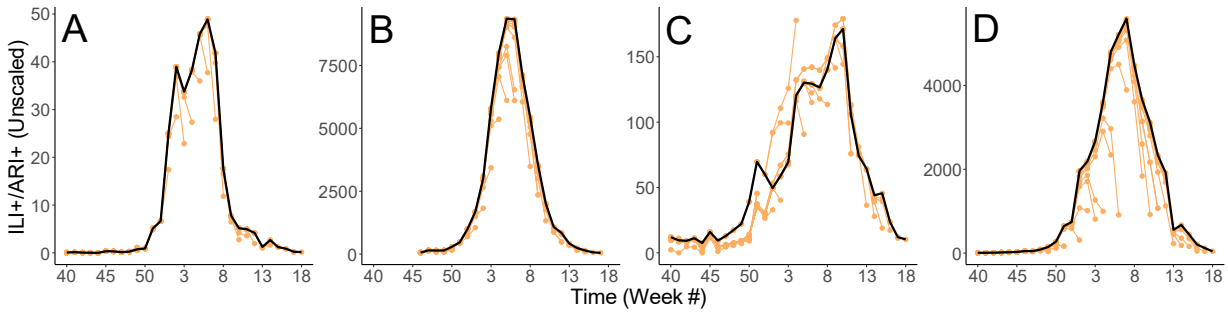


Figure 3. Representative examples of data updates over time. Black lines represent the final unscaled data values, downloaded after the end of the season (week 39). Orange points show the data downloaded each week in real-time. Data are displayed for (A) Ireland, 2018-19; (B) Italy, 2018-19; (C) the Netherlands, 2017-18; and (D) Norway, 2017-18.

Even when data are reported in real-time, values are sometimes updated at later weeks as more information becomes available (Figure 3). Throughout a given season, the proportion of real-time data values that were later updated ranges from 0 to 1 (mean = 0.414; median = 0.199), although we note that many times these updates were relatively small in magnitude (Figure 3A-C), and seven countries had no updates for either season. Typically, the real-time values underestimated the end-of-season values, although this was not always the case (Figure 3C). Furthermore, real-time values were sometimes adjusted to their final values as soon as the next week (Figure 3A), but could also be adjusted over the course of several weeks (Figure 3C-D). As with missingness, there was no significant difference in update rates by season (Kruskal-Wallis test, $p > 0.85$).

Real-time forecast accuracy

In general, we found that skillful real-time forecasts of influenza activity for the 2017-18 and 2018-19 seasons were possible (Figure 4). We note here that, because data are reported with a one-week lag, a predicted lead week of -1 in reality represents a forecast generated at the predicted peak. Forecast accuracy was higher during the 2018-19 season than the 2017-18 season for both peak timing and peak intensity at almost all predicted lead weeks. The proportion of forecasts accurately predicting peak timing consistently exceeded 50% starting at predicted lead week -1 for the 2017-18 season and at week -3 for the 2018-19 season; accuracy exceeded 75% beginning at lead week 3 for 2017-18 and week -1 for 2018-19. For peak intensity, forecast accuracy first exceeded 50% at a predicted lead week of 0 for 2017-18 and -1 for 2018-19, and also exceeded 75% starting at lead week -1 during the 2018-19 season; accuracy never exceeded 75% during the 2017-18 season. The difference in accuracy by season was found to be statistically significant overall (peak timing: aOR = 2.404, 95% CI: 1.198–4.813; peak intensity: aOR = 2.295, 95% CI: 1.243–4.199), but not before the predicted peak. Similar patterns emerge when results are assessed by observed lead week (S1 Figure).

We previously observed that retrospective forecasts were significantly less accurate for countries reporting ARI data than for countries reporting ILI data (see Chapter 2). Here, we find that countries reporting ILI data yielded more accurate forecasts for peak timing but not for peak intensity, and that countries reporting ARI data actually appeared to produce more accurate forecasts of peak intensity at lead weeks -4 through -2 (S2 Figure). While the difference in peak timing accuracy (aOR = 0.413, 95% CI: 0.189–0.896) was significant before the predicted peak, the difference in peak intensity accuracy was not. No significant patterns were observed by region (S3 Figure). Adjusted odds ratios for all comparisons can be found in S2 Table.

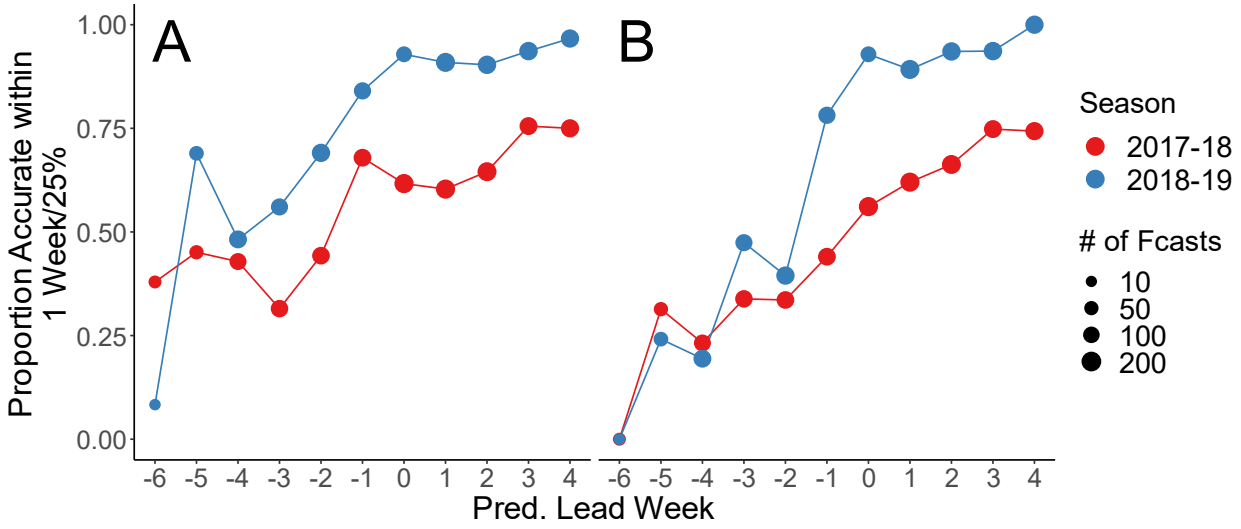


Figure 4. Real-time forecast accuracy by predicted lead week. Forecast accuracy is shown for (A) peak timing and (B) peak intensity. Results for the 2017-18 season are shown in red, and results for 2018-19 are in blue. The size of the points represents the number of forecasts generated for that predicted lead week.

Compared to historical expectance: We compared our real-time forecasts to forecasts produced using the method of analogues, a non-mechanistic method that produces forecasts based on historical patterns (Viboud et al., 2003). We found that our methods and the method of analogues perform similarly for the 2017-18 season, but that our methods outperformed the method of analogues for the 2018-19 season beginning at predicted lead week -1 for peak timing and -3 for peak intensity (S4 Figure).

Compared to retrospective forecast accuracy: Forecasts produced using complete and updated data from the end of the season did not appear more accurate than real-time forecasts, although the number of forecasts produced at each predicted lead week was slightly higher for retrospective than real-time forecasts (Figure 5). When assessed by observed lead week, including only those forecast pairs produced both in real-time and retrospectively, retrospective forecasts performed significantly better than real-time forecasts at lead weeks -5 and -1 for peak

timing, and at lead weeks -2 and -1 for peak intensity (Table 1). Note that we used a Bonferroni-corrected p-value cutoff of 0.05 / 11 = 0.0045.

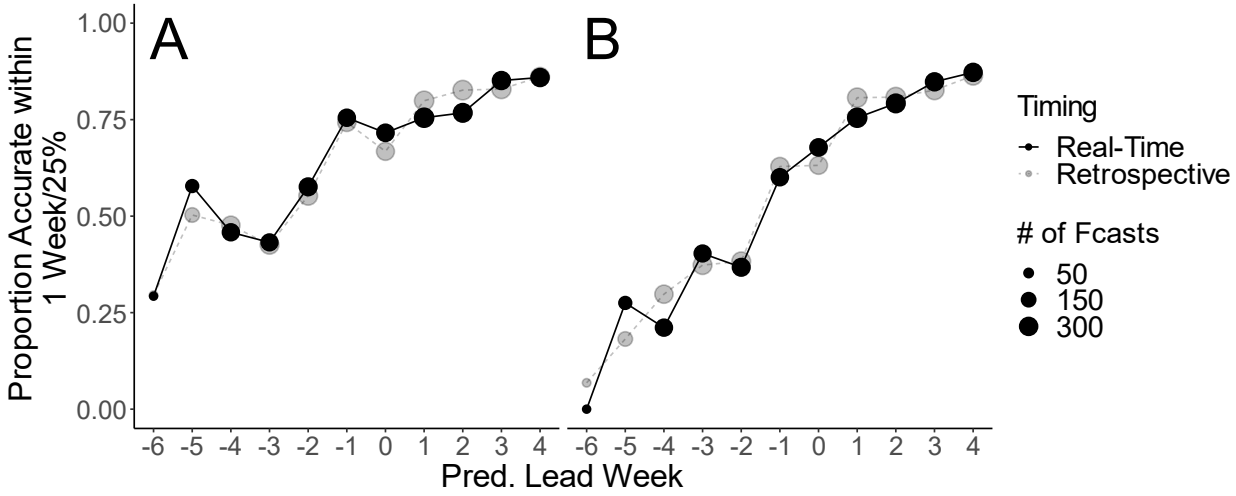


Figure 5. Real-time versus retrospective forecast accuracy. Forecast accuracy for (A) peak timing and (B) peak intensity is shown for both real-time (in black, with solid lines) and retrospective (in gray, with dotted lines) forecasts. The size of each point represents the number of forecasts generated.

Table 1. Accuracy of retrospective vs. real-time forecasts by observed lead week.

Obs. Lead Week:		-6	-5	-4	-3	-2	-1	0	1	2	3	4
Timing	Retro.	11.3%	24.7%	31.7%	48.3%	54.6%	61.3%	67.7%	79.3%	90.7%	90.5%	94.0%
	RT	11.3%	20.0%	28.3%	42.8%	50.8%	56.5%	66.8%	81.3%	91.8%	91.3%	93.6%
	Sig.		**				*					
Intensity	Retro.	9.2%	10.6%	15.7%	25.5%	43.1%	51.9%	71.0%	86.3%	90.7%	90.9%	90.2%
	RT	5.6%	10.6%	16.7%	20.7%	32.2%	45.2%	70.0%	87.0%	91.4%	89.1%	90.6%
	Sig.					**	*					

* p<0.0045
**p<0.001

Real-time forecast calibration

In addition to being accurate, a skillful forecast should also be well-calibrated. That is, ensemble forecasts with less variability between ensemble members should indicate higher certainty. Here, we define a well-calibrated model as one in which the observed peak timing or intensity falls within the *n*th prediction interval, as delineated by ensemble variance, *n*% of the time. Figure 6 compares forecast calibration before the predicted peak for real-time and

retrospective forecasts. We found that both real-time and retrospective forecasts were relatively well-calibrated, and that retrospective forecasts did not appear to be any better calibrated than real-time forecasts.

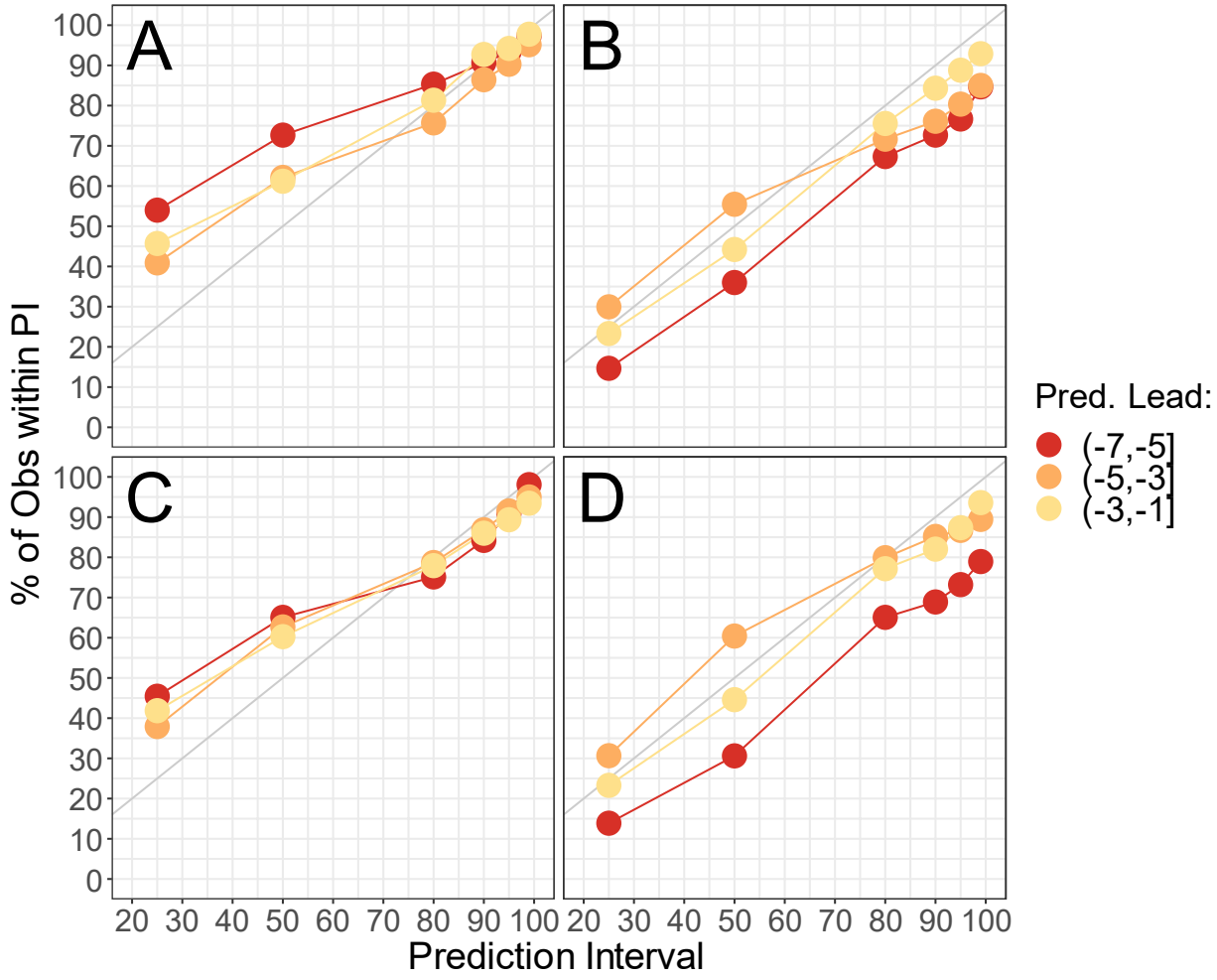


Figure 6. Real-time and retrospective forecast calibration. Calibration is shown as the percentage of observations (on the y-axis) that fell within various forecast prediction intervals (x-axis), where prediction intervals are calculated based on the distribution of ensemble members, for a range of predicted lead weeks. (A) Real-time forecasts, peak timing; (B) Real-time forecasts, peak intensity; (C) Retrospective forecasts, peak timing; (D) Retrospective forecasts, peak intensity

Discussion

Forecasts are a potentially important tool for public health response against seasonal influenza outbreaks, because high-quality forecasts could provide the public health and medical

communities with several weeks of additional preparation time. Here, we describe the results from two seasons of real-time forecasting of influenza outbreaks in Europe, using data sources and models previously shown to be capable of producing skillful retrospective forecasts in the same countries. We find that real-time forecasting accuracy was broadly consistent with these past retrospective results (Chapter 2). Forecast accuracy was significantly higher for the 2018-19 season than for the 2017-18 season, and only the 2018-19 season was consistently more accurately forecasted using our model-inference system than by the method of analogues, a method based on historical expectance. While it is difficult to say what exactly drove this difference, it is possible that the extended duration of the 2017-18 season made it more difficult to forecast using a simple compartmental model. Consistent with our past findings, we find that real-time forecast accuracy for peak timing was higher for countries reporting ILI data than for countries reporting ARI data, at least before the predicted peak. As ILI data are more specific and tend to be less noisy, this finding is not surprising.

It was encouraging to see that, despite substantial late reporting and data updating over time, real-time forecasts were not significantly less accurate than retrospective forecasts at most timepoints, nor were they less well-calibrated. This suggests that small errors in data reported in real-time are unlikely to greatly impact forecast quality. Moss et al. (2018) compared real-time and retrospective forecasts for a season with unusually long delays in reporting and found retrospective forecasts to be noticeably more accurate, suggesting that this finding may not hold in the case of larger updates. Future work should explore the incorporation of backcasts, which attempt to correct for these inaccuracies throughout the existing data, as such work has previously been shown to improve forecast accuracy (Kandula et al., 2019).

Although data reported in real-time were generally not updated substantially throughout the season, countries reported no real-time data for roughly 30% of weeks, on average. This lack of timely reporting led to significant missed opportunities, as no real-time forecasts could be generated for these countries and weeks. Timely reporting was particularly rare over the winter holidays (weeks 51 and 52). While modelers need to improve forecast model systems in order to produce more skillful real-time forecasts, the importance of delivering available, timely and accurate data streams must continue to be emphasized. The development of partnerships between modelers and surveillance experts, as cultivated by Moss et al. (2018), and as further discussed below, could be beneficial here: insight from public health professionals can guide model adjustments to help compensate for reporting delays (Moss et al., 2018, 2019), while public health practitioners exposed to the potential benefits of forecasting may be more likely to push for more timely and accurate reporting.

We have discussed the primary limitations of the FluNet and FluID data, including lack of specificity and denominator data, in Chapter 2. Here, we focus on limitations with particular relevance for real-time forecasting. First, we reiterate that data are reported with a one-week lag, meaning that the forecasts are not truly produced in "real time," and that one-week-ahead forecasts are actually "nowcasts" (Lampos et al., 2015). This is a common issue with influenza surveillance data (Biggerstaff et al., 2018; Kandula et al., 2019). Future work should consider alternative data streams and nowcasting methods, i.e. methods for estimating the current number of cases, which are unobserved due to the one-week lag in reporting, which may improve forecast accuracy (Kandula et al., 2017).

Secondly, because we have results for only two influenza seasons, the practical utility of these forecasts should be judged as preliminary. However, we note that forecasts of peak

intensity for both seasons outperformed the method of analogues, a robust method based on historical expectance, in the three weeks leading up to the predicted peak. Thus, although peak intensity forecast accuracy was generally below 50% before the peak, our forecasts still provide more information about the magnitude of an upcoming peak than can be gleaned using historical data alone, and can therefore contribute to better-informed decisions during the weeks leading up to the peak. Additionally, forecasts for the 2018-19 season were very accurate after the predicted peak had passed. If this holds for future seasons, real-time forecasts would allow for higher certainty when declaring that an outbreak is in decline. Overall, it appears that real-time forecasts could begin to be operationalized for countries where forecasts outperform historical expectance, but that modelers and practitioners should proceed with caution until more evidence from future seasons is amassed. Finally, we note that country-level forecasts may be of limited use in particularly large countries. Regional or city-level data could allow for the development of more local forecasts, which may yield more practical results.

Finally, we note that accurate and well-calibrated models face several barriers to use. Many public health practitioners lack knowledge of or trust in mathematical models (Driedger et al., 2014; Muscatello et al., 2017), and convincing them to use models in developing outbreak responses will likely require improved communication between modelers and practitioners, as well as longstanding partnerships (Driedger et al., 2014; Metcalf et al., 2015; Moghadas et al., 2009, 2015; Moss et al., 2018). That said, the use of models to inform public health decisions regarding seasonal influenza is not unprecedented. The United Kingdom's Joint Committee on Vaccination and Immunisation (*JCVI*, n.d.), for example, takes results from mathematical modeling efforts into account when deciding on seasonal influenza vaccination strategies (Joint Committee on Vaccination and Immunisation, 2020). Models also helped to inform the

implementation of various control measures during the 2009 H1N1 pandemic (Lee et al., 2013), and models were used as early as the 1970s to anticipate influenza activity in the Soviet Union (Ivanov & Leonenko, 2017; Longini, 1988). Promisingly, both the WHO and the CDC show interest in the further development and use of forecasts to be used in the public health response to influenza (Biggerstaff et al., 2018, 2019).

Conclusions

We have shown that skillful real-time forecasts of influenza activity can be produced for several countries, and that current levels of late and inaccurate reporting do not appear to substantially reduce forecast quality. The lack of timely reporting is of particular importance, as no new forecast can be produced for a given week if new data are not reported, representing a substantial missed opportunity. On the modeling side, the usefulness of nowcasting methods in improving forecast accuracy should be considered. Future work should also explore the potential for forecasting on smaller spatial scales, as well as optimal ways of communicating forecast certainty to non-modelers. Finally, while the recent interest of organizations such as the WHO and CDC in forecasting is encouraging, efforts to build long-lasting relationships between modelers and public health practitioners remain important. These relationships not only help to build trust in models among non-modelers, but also provide modelers with valuable insight into the data we use.

References

Anderson, J. L. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;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., Velardi, P., Vespignani, A., & Finelli, L. (2016). Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infectious Diseases*, *16*. https://doi.org/10.1186/s12879-016-1669-x

Biggerstaff, M., Dahlgren, F. S., Fitzner, J., George, D., Hammond, A., Hall, I., Haw, D., Imai, N., Johansson, M. A., Kramer, S., McCaw, J. M., Moss, R., Pebody, R., Read, J. M., Reed, C., Reich, N. G., Riley, S., Vandemaele, K., Viboud, C., & Wu, J. T. (2019). Coordinating the real-time use of global influenza activity data for better public health planning. *Influenza and Other Respiratory Viruses*, *14*(2), 105–110. https://doi.org/10.1111/irv.12705

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. (2018). Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*. https://doi.org/10.1016/j.epidem.2018.02.003

Bloom-Feshbach, K., Alonso, W. J., Charu, V., Tamerius, J., Simonsen, L., Miller, M. A., & Viboud, C. (2013). Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PLOS ONE*, *8*(2), e54445. https://doi.org/10.1371/journal.pone.0054445

Columbia University Mailman School of Public Health. (n.d.). *Influenza Observations and Forecast.* Columbia Prediction of Infectious Diseases. https://cpid.iri.columbia.edu/

Driedger, S. M., Cooper, E. J., & Moghadas, S. M. (2014). Developing model-based public health policy through knowledge translation: The need for a 'Communities of Practice.' *Public Health*, *128*(6), 561–567. https://doi.org/10.1016/j.puhe.2013.10.009

Farrow, D. C., Brooks, L. C., Hyun, S., Tibshirani, R. J., Burke, D. S., & Rosenfeld, R. (2017). A human judgment approach to epidemiological forecasting. *PLoS Computational Biology*, *13*(3). https://doi.org/10.1371/journal.pcbi.1005248

Hammond, A., Hundal, K., Laurenson-Schafer, H., Cozza, V., Maharjan, B., Fitzner, J., Samaan, M., Vandemaele, K., & Wenqing, Z. (2019). Review of the 2018–2019 influenza season in the northern hemisphere. *Weekly Epidemiological Record*, *94*(32), 345–364.

Hammond, A., Laurenson-Schafer, H., Marsland, M., Besselaar, T., Fitzner, J., Vandemaele, K., & Zhang, W. (2018). Review of the 2017-2018 influenza season in the northern hemisphere. *Weekly Epidemiological Record*, *93*(34), 429–444.

Hickmann, K. S., Fairchild, G., Priedhorsky, R., Generous, N., Hyman, J. M., Deshpande, A., & Del Valle, S. Y. (2015). Forecasting the 2013–2014 Influenza Season Using Wikipedia. *PLoS Computational Biology*, *11*(5). https://doi.org/10.1371/journal.pcbi.1004239

Ivanov, S. V., & Leonenko, V. N. (2017). Prediction of influenza peaks in Russian cities: Comparing the accuracy of two SEIR models. *Mathematical Biosciences and Engineering*, *15*(1), 209–232. https://doi.org/10.3934/mbe.2018009

Joint Committee on Vaccination and Immunisation. (n.d.). GOV.UK. https://www.gov.uk/government/groups/joint-committee-on-vaccination-and-immunisation

Joint Committee on Vaccination and Immunisation. (2020). *Minute of the meeting held on 04 and 05 February 2020*. https://app.box.com/s/iddfb4ppwkmtjusir2tc/file/636396626894

Kandula, S., Hsu, D., & Shaman, J. (2017). Subregional Nowcasts of Seasonal Influenza Using Search Trends. *Journal of Medical Internet Research*, *19*(11), e370. https://doi.org/10.2196/jmir.7486

Kandula, S., Pei, S., & Shaman, J. (2019). Improved forecasts of influenza-associated hospitalization rates with Google Search Trends. *Journal of The Royal Society Interface*, *16*(155), 20190080. https://doi.org/10.1098/rsif.2019.0080

Kramer, S. C., & Shaman, J. (2019). Development and validation of influenza forecasting for 64 temperate and tropical countries. *PLOS Computational Biology*, *15*(2), e1006742. https://doi.org/10.1371/journal.pcbi.1006742

Lampos, V., Miller, A. C., Crossan, S., & Stefansen, C. (2015). Advances in nowcasting influenza-like illness rates using search query logs. *Scientific Reports*, *5*(1), 12760. https://doi.org/10.1038/srep12760

Lee, B. Y., Haidari, L. A., & Lee, M. S. (2013). Modelling during an emergency: The 2009 H1N1 influenza pandemic. *Clinical Microbiology and Infection*, *19*(11), 1014–1022. https://doi.org/10.1111/1469-0691.12284

Longini, I. M. (1988). A mathematical model for predicting the geographic spread of new infectious agents. *Mathematical Biosciences*, *90*(1–2), 367–383.

Metcalf, C. J. E., Edmunds, W. J., & Lessler, J. (2015). Six challenges in modelling for public health policy. *Epidemics*, *10*, 93–96. https://doi.org/10.1016/j.epidem.2014.08.008

Moghadas, S. M., Haworth-Brockman, M., Isfeld-Kiely, H., & Kettner, J. (2015). Improving Public Health Policy through Infection Transmission Modelling: Guidelines for Creating a Community of Practice. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *26*(4), 191–195. https://doi.org/10.1155/2015/274569

Moghadas, S. M., Pizzi, N. J., Wu, J., & Yan, P. (2009). Managing public health crises: The role of models in pandemic preparedness. *Influenza and Other Respiratory Viruses*, *3*(2), 75–79. https://doi.org/10.1111/j.1750-2659.2009.00081.x

Moss, R., Fielding, J. E., Franklin, L. J., Stephens, N., McVernon, J., Dawson, P., & McCaw, J. M. (2018). Epidemic forecasts as a tool for public health: Interpretation and (re)calibration. *Australian and New Zealand Journal of Public Health*, *42*(1), 69–76. https://doi.org/10.1111/1753-6405.12750

Moss, R., Zarebski, A., Dawson, P., & McCAW, J. M. (2017). Retrospective forecasting of the 2010–2014 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology & Infection*, *145*(1), 156–169. https://doi.org/10.1017/S0950268816002053

Moss, R., Zarebski, A. E., Dawson, P., Franklin, L. J., Birrell, F. A., & McCaw, J. M. (2019). Anatomy of a seasonal influenza epidemic forecast. *Communicable Diseases Intelligence*, *43*. https://doi.org/10.33321/cdi.2019.43.7

Muscatello, D. J., Chughtai, A. A., Heywood, A., Gardner, L. M., Heslop, D. J., & MacIntyre, C. R. (2017). Translation of Real-Time Infectious Disease Modeling into Routine Public Health Practice. *Emerging Infectious Diseases*, *23*(5). https://doi.org/10.3201/eid2305.161720

Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., & Marathe, M. V. (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses*, *8*(3), 309–316. https://doi.org/10.1111/irv.12226

Ong, J. B. S., Chen, M. I.-C., Cook, A. R., Lee, H. C., Lee, V. J., Lin, R. T. P., Tambyah, P. A., & Goh, L. G. (2010). Real-Time Epidemic Monitoring and Forecasting of H1N1-2009 Using Influenza-Like Illness from General Practice and Family Doctor Clinics in Singapore. *PLoS ONE*, *5*(4), e10036. https://doi.org/10.1371/journal.pone.0010036

Pei, S., Kandula, S., Yang, W., & Shaman, J. (2018). Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences*, 201708856. https://doi.org/10.1073/pnas.1708856115

Rodell, M. (n.d.). *LDAS | Land Data Assimilation Systems* [Text.Journal]. https://ldas.gsfc.nasa.gov/gldas/GLDASgoals.php

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, *109*(50), 20425–20430. https://doi.org/10.1073/pnas.1208772109

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3837

Viboud, C., Boëlle, P.-Y., Carrat, F., Valleron, A.-J., & Flahault, A. (2003). Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, *158*(10), 996–1006.

WHO. (n.d.-a). *WHO | FluID - a global influenza epidemiological data sharing platform*. WHO. https://www.who.int/influenza/surveillance_monitoring/fluid/en/

WHO. (n.d.-b). *WHO | FluNet*. WHO. https://www.who.int/influenza/gisrs_laboratory/flunet/en/

Yang, W., Cowling, B. J., Lau, E. H. Y., & Shaman, J. (2015). Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology*, *11*(7), e1004383. https://doi.org/10.1371/journal.pcbi.1004383

Supplementary Tables

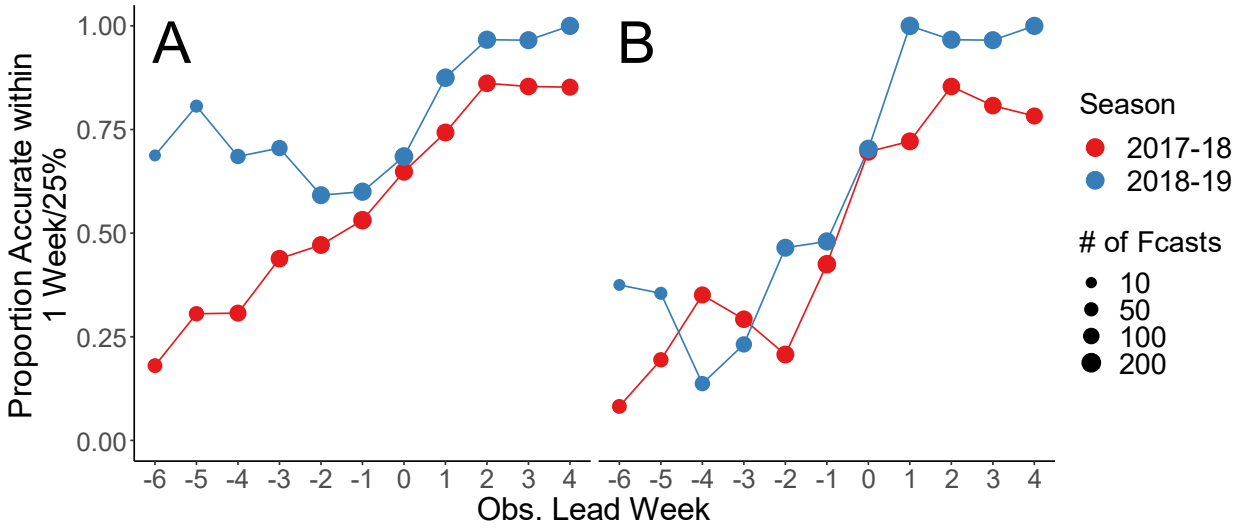
S1 Table. Countries used for real-time forecasting by region, data type, and season-specific scaling. Values that change for the 2018-19 season are bolded and italicized.

Country	Region	Data Type	Scaling (17-18)	Scaling (18-19)
Austria	Southwest Europe	ILI	14.0	14.0
Belgium	Southwest Europe	ILI	8.5	8.5
Croatia	Southwest Europe	ILI	1.4	1.4
France	Southwest Europe	ILI	0.03	0.03
Germany	Southwest Europe	ARI	0.25	0.25
Greece	Southwest Europe	ILI	4.0	4.0
Italy	Southwest Europe	ILI	1.0	***0.75***
Luxembourg	Southwest Europe	ARI	32.0	32.0
Netherlands	Southwest Europe	ILI	31.0	***26.0***
Portugal	Southwest Europe	ILI	245.0	245.0
Serbia	Southwest Europe	ILI	0.5	0.5
Slovenia	Southwest Europe	ARI	5.8	5.8
Spain	Southwest Europe	ILI	2.0	2.0
Belarus	Eastern Europe	ARI	0.2	0.2
Bulgaria	Eastern Europe	ARI	1.25	1.25
Czechia	Eastern Europe	ILI	0.8	0.80
Georgia	Eastern Europe	ILI	15.0	15.0
Hungary	Eastern Europe	ILI	1.0	1.0
Israel	Eastern Europe	ILI	3.5	3.5
Kazakhstan	Eastern Europe	ARI	0.35	0.35
Kyrgyzstan	Eastern Europe	ARI	0.63	0.63
Poland	Eastern Europe	ILI	1.3	***1.0***
Republic of Moldova	Eastern Europe	ARI	2.25	2.25
Romania	Eastern Europe	ILI	39.0	39.0
Russian Federation	Eastern Europe	ARI	0.02	0.02
Slovakia	Eastern Europe	ILI	0.57	0.57
Turkey	Eastern Europe	ILI	3.5	3.5
Ukraine	Eastern Europe	ARI	0.03	0.03
Uzbekistan	Eastern Europe	ARI	68.0	***54.0***
Denmark	Northern Europe	ILI	18.0	18.0
Estonia	Northern Europe	ARI	1.5	1.5
Finland	Northern Europe	ILI	14.0	***7.0***
Iceland	Northern Europe	ILI	57.0	57.0
Ireland	Northern Europe	ILI	45.0	45.0
Latvia	Northern Europe	ARI	4.4	4.4
Lithuania	Northern Europe	ILI	1.3	***1.2***
Norway	Northern Europe	ILI	1.65	***1.1***

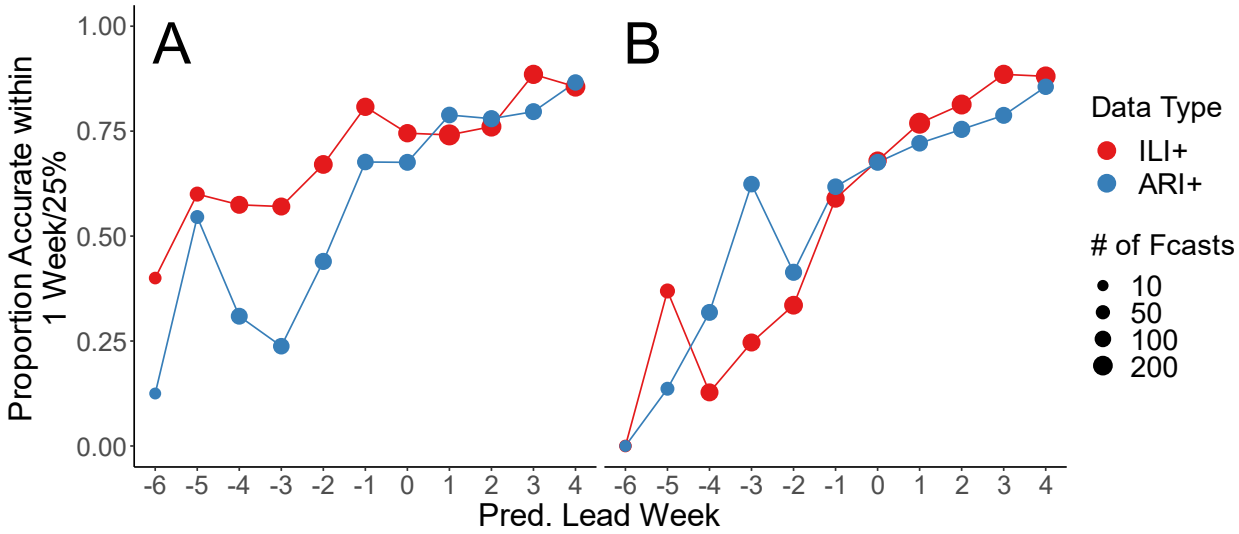
S2 Table. Real-time forecast accuracy overall, and before the predicted peak, by season, data type, and region. Seasonal differences were assessed using the 2017-18 season as the reference level, while ILI+ was used as the reference level when comparing data types. Southwest Europe was used as the reference level for the analysis by region, as indicated. Cells shaded green indicate improved accuracy over the reference level, whereas cells shaded red indicate a reduction in accuracy.

Variable		Peak Timing aOR (95% CI)		Peak Intensity aOR (95% CI)	
		Overall	Before Peak	Overall	Before Peak
Season	2018-19	2.404 (1.198, 4.813)	1.789 (0.836, 3.836)	2.295 (1.243, 4.199)	1.501 (0.766, 2.964)
Data Type	ARI+	0.602 (0.301, 1.206)	0.413 (0.189, 0.896)	1.255 (0.648, 2.406)	1.841 (0.929, 3.631)
Region	Southwest Europe	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
	Eastern Europe	0.962 (0.437, 2.094)	0.860 (0.369, 2.007)	1.187 (0.572, 2.422)	1.012 (0.479, 2.114)
	Northern Europe	1.404 (0.455, 4.355)	1.308 (0.450, 3.782)	0.756 (0.318, 1.795)	0.560 (0.222, 1.416)

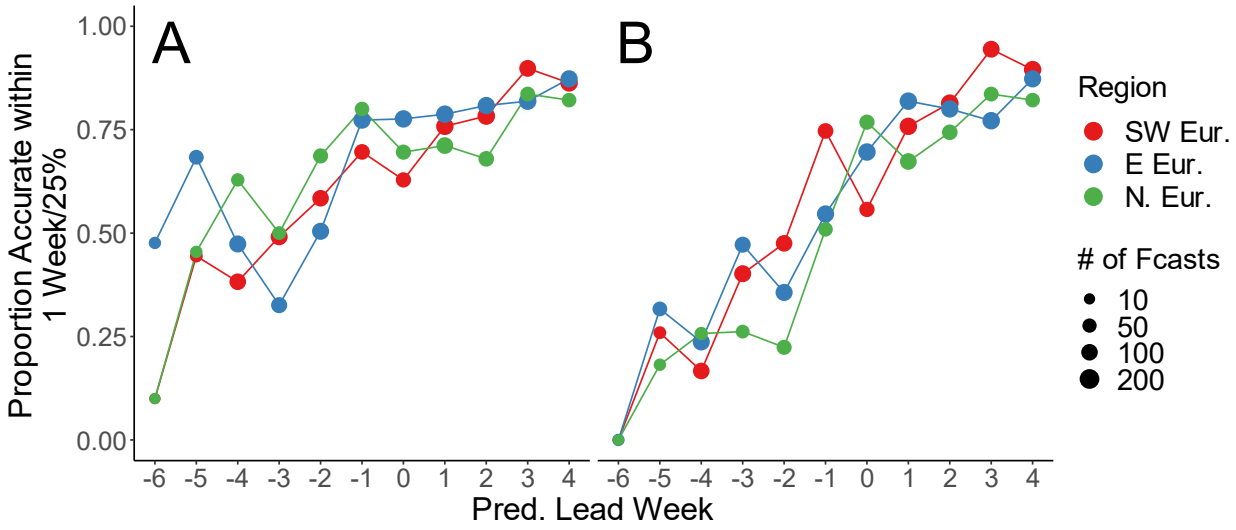
Supplementary Figures



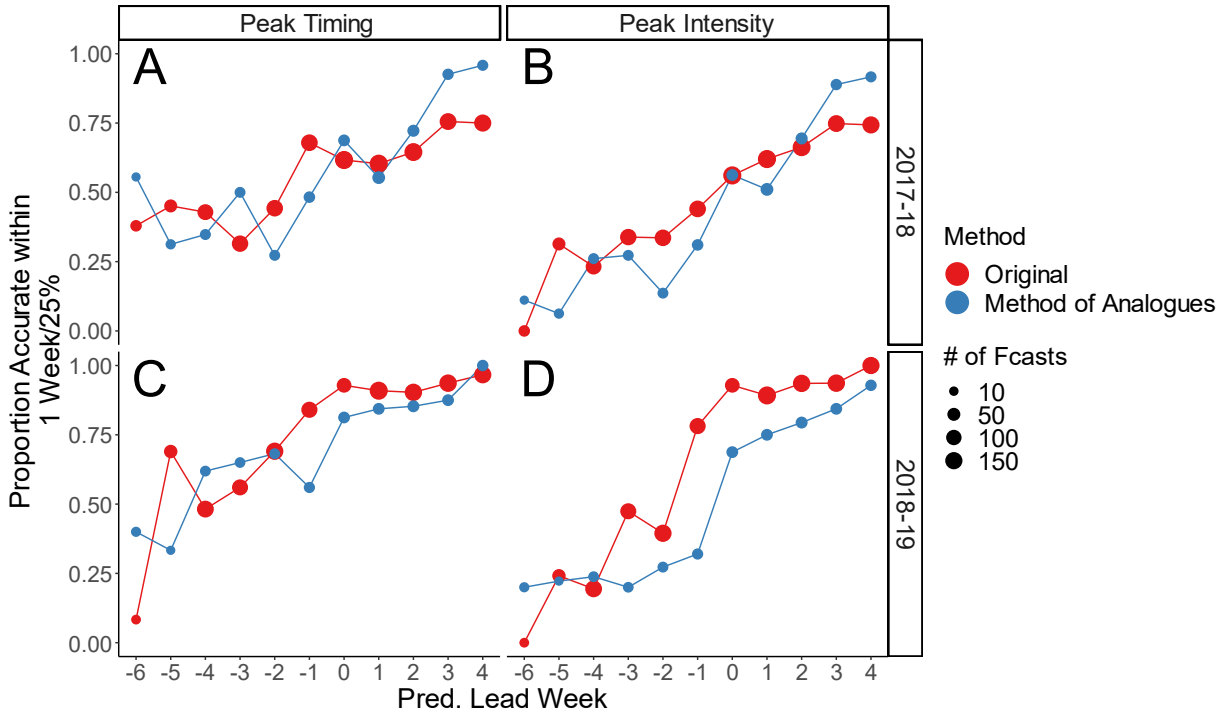
S1 Figure. Real-time forecast accuracy by observed lead week. Forecast accuracy for (A) peak timing and (B) peak intensity are shown for the 2017-18 (red) and 2018-19 (blue) seasons. The number of forecasts generated at each lead week is represented by point size.



S2 Figure. Real-time forecast accuracy by data type. (A) Peak timing and (B) peak intensity forecast accuracy are compared for countries reporting ILI (red) and ARI (blue) data. Point size indicates the number of forecasts generated at each lead week.



S3 Figure. Real-time forecast accuracy by region. The accuracy of (A) peak timing and (B) peak intensity forecasts is shown by region. Point size indicates the number of forecasts produced.



S4 Figure. Comparing forecast accuracy using the method of analogues. Peak timing (A and C) and peak intensity (B and D) forecast accuracy is shown separately for the 2017-18 season (A and B) and the 2018-19 season (C and D). Forecasts using the methods described in the main text are shown in red, while results obtained using the method of analogues are shown in blue. The size of the points shows how many forecasts were generated at each predicted lead week.

Chapter 5

Assessing the Use of Influenza Forecasts and Epidemiological Modeling in Public Health Decision Making in the United States

Sarah C. Kramer[1], Colin Doms[2], and Jeffrey Shaman[1]

Affiliations:
1. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, 722 W 168th Street, New York, New York, 10032, United States of America
2. Quantitative Methods in the Social Sciences, Graduate School of Arts and Sciences, Columbia University, 109 Low Memorial Library, MC 4306, 535 W 116th Street, New York, New York, 10027, United States of America

Acknowledgments:
The authors would like to thank ASTHO, CSTE, and NACCHO for their cooperation and willingness to survey their members.

Published: Doms, C., Kramer, S.C., & Shaman, J. 2018. Assessing the Use of Influenza Forecasts and Epidemiological Modeling in Public Health Decision Making. *Sci Rep, 8*(1).

Abstract

Although forecasts and other mathematical models have the potential to play an important role in mitigating the impact of infectious disease outbreaks, the extent to which these tools are used in public health decision making in the United States is unclear. Throughout 2015, we invited public health practitioners belonging to three national public health organizations to complete a cross-sectional survey containing questions on model awareness, model use, and communication with modelers. Of 39 respondents, 46.15% used models in their work, and 20.51% reported direct communication with those who create models. Over half (64.10%) were aware that influenza forecasts exist. The need for improved communication between practitioners and modelers was overwhelmingly endorsed, with over 50% of participants indicating the need for models more relevant to public health questions, increased frequency of telecommunication, and more plain language in discussing models. Model use for public health decision making must be improved if models are to reach their full potential as public health tools. Increased quality and frequency of communication between practitioners and modelers could be particularly useful in achieving this goal. It is important that improvements be made now, rather than waiting for the next public health crisis to occur.

Background

Numerical forecasting—the computational real-time generation of calibrated predictions on time scales allowing application and validation—has a long history of use in the fields of weather and climate (Gneiting & Raftery, 2005; Zebiak et al., 2015; Zebiak & Cane, 1987). In recent decades, numerical forecasts have been developed for and applied to a number of new industries and disciplines, including agriculture (FAO, 2016; Newlands et al., 2014), air quality (Debry & Mallet, 2014; Gaubert et al., 2014), consumer activity (Chen & Lu, 2017; Choi et al., 2014; Mccarthy et al., 2006), fiscal policy (Sun, 2014), and political elections (Berg et al., 2008). These forecasts allow stakeholders to prepare for predicted future events and to respond accordingly. For example, forecasts of crop yields help governments decide whether food must be imported to meet population needs, and inform decisions concerning the receipt of emergency food aid (Newlands et al., 2014). Meanwhile, many companies use sales forecasting when deciding how much of a product to stock in order to maximize profits (Chen & Lu, 2017). In public health, forecasting methods have been developed using mathematical models and Bayesian inference methods and used to predict the growth and spread of infectious diseases such as influenza (Hickmann et al., 2015; Ong et al., 2010; Shaman et al., 2013; Shaman & Karspeck, 2012; Viboud et al., 2003; Yang et al., 2015), dengue (Adde et al., 2016; Reich et al., 2016; Shi et al., 2015), Ebola (Camacho et al., 2015; Meltzer et al., 2014; Shaman et al., 2014), and, most recently, Zika (Chowell et al., 2016; Huff et al., 2016).

In the United States, influenza is estimated to kill tens of thousands of people and cost over $87 billion each year (Molinari et al., 2007). Several research groups, including ours, have developed forecasts of influenza incidence in the United States (CDC, n.d.). These forecasts estimate future incidence levels for a developing influenza outbreak with particular focus on

metrics such as when the outbreak will be most severe or how many cases will occur during the most severe week of the outbreak. In our own efforts, forecasts have been generated for municipalities and states throughout the US, as well as for several European countries, and operationalized for real-time delivery over an online portal (Columbia University Mailman School of Public Health, n.d.). These quantitative forecasts are updated weekly during the flu season and have the potential to reduce morbidity, mortality, and healthcare spending by influencing decision making and resource allocation among healthcare providers, public health practitioners, and the general public alike. For example, hospitals may use the forecast peak timing of an influenza outbreak to prepare for an influx of patients, and the public may be more motivated to practice proper hand hygiene when high influenza incidence is predicted. However, these benefits will only be fully realized if public health practitioners are aware of this work and use these findings in decision making.

Research on the extent to which public health practitioners utilize mathematical models is limited. Indeed, to our knowledge, no existing studies assess the use of mathematical models in public health decision making in the US. Driedger et al. (2014) interviewed four public health practitioners and four mathematical modelers in order to assess the integration of modeling in decision making during the 2009 influenza pandemic in Canada. They concluded that improved communication between practitioners and modelers was needed. Specifically, they found that practitioners desired greater clarity in model interpretation, and modelers wanted a better understanding of the questions practitioners needed modeled. Both groups expressed the need for longstanding partnerships in order to increase efficiency, understanding, and trust between the two groups. An earlier Canadian study also found need for more and better communication between practitioners and modelers (Moghadas et al., 2009). Most recently, Moss et al. (2018)

shared weekly forecasts of influenza activity in Melbourne, Australia with the local health department, and updated their forecasts based on insights from the practitioners there. They report that these collaborations were instrumental in improving forecast accuracy.

Here, we addressed these issues using a different approach. We employed a short survey to assess the extent to which US public health professionals are aware of and use mathematical models, including influenza forecasts, in making decisions on the job. Through this preliminary effort, we seek to build the evidence base describing the integration of numerical epidemiological modeling, including seasonal influenza forecast, and public health decision making.

Methods

Participants

We recruited survey participants via email through contacts at three U.S. public health organizations: the Association of State and Territorial Health Officials (ASTHO), the Council of State and Territorial Epidemiologists (CSTE), and the National Association of County and City Health Officials (NACCHO). Although we do not know how many practitioners ultimately received a link to our survey, these organizations represent a large number of employees in the fields of public health, epidemiology, and influenza control across the US, ensuring that our survey was sent to a representative sample of US public health practitioners.

Materials

We designed a survey containing 25 multiple-choice and Likert scale questions (see Supplementary Information). The survey included questions on basic demographics, awareness

of influenza forecasts, whether the respondents used epidemiological models in their work, and whether they applied model results to public health decision making. Participants were also asked if they communicated with modelers, and how such communication could be improved. Finally, we inquired about personal use of influenza vaccination for the current and previous seasons. This work was approved by and performed under Columbia University Medical Center IRB (approval number CUMC IRB-AAAO9952). The IRB-approved survey was distributed online through SurveyMonkey, and informed consent was acquired through a checkbox on the survey's first page. All results were de-identified.

Procedure

Participants were recruited through broadcast emails to the members of each of the three organizations. We collected responses over roughly a six-month period. Most of the responses from one organization were collected during March and April 2015, and other responses were completed during September 2015. The difference in timing was due to differing availability to contact their members. In addition, in August 2015 we changed the word 'survey' to 'assessment' in order to comply with a request from one organization and gather more responses. Thus, a majority of participants saw 'assessment', although we believe this wording change had little effect, if any, on the results.

Results

Data

A total of 51 individuals responded to the survey, 42 (82.4%) of whom indicated employment in a public health field. Because we are primarily interested in awareness and use of

models among public health practitioners, we restricted our analysis to these individuals. Furthermore, we removed three other participants whose responses were inconsistent; specifically, two individuals reported a frequency of model use while simultaneously reporting that they did not use models in their work, and one participant reported acquiring influenza data from both Columbia University and none of the sources listed on the survey. This left us with data on 39 participants. All 39 participants reported that their work-related responsibilities included planning for and dealing with influenza outbreaks. The majority of respondents (38, 97.4%) worked for the government, and one worked for an NGO.

Demographics

Demographic information is summarized in Table 1. Briefly, the majority of respondents (22, 56.4%) were between the ages of 30 and 49. Years in public health was fairly evenly distributed, with the largest group being those who had been in the field for 4-6 years (13, 33.3%). Two-thirds of respondents (26, 66.7%) reported being female, and most (33, 84.6%) had at least a graduate degree. Respondents were spread geographically across 35 states and territories. Regional totals are based on

Table 1. Demographic Characteristics of 39 Public Health Practitioners Surveyed Concerning Awareness and Use of Mathematical Models.

Gender	
Female	26 (66.7%)
Male	12 (30.8%)
Age	
18-29	9 (23.1%)
30-49	22 (56.4%)
50-64	5 (12.8%)
65+	1 (2.6%)
Degree obtained	
Bachelor's degree	6 (15.4%)
Graduate degree	33 (84.6%)
Years in public health	
0-3 years	5 (12.8%)
4-6 years	13 (33.3%)
7-10 years	6 (15.4%)
11-15 years	8 (20.5%)
16+ years	6 (15.4%)
Region	
West	10 (25.6%)
South	10 (25.6%)
Northeast	7 (17.9%)
Midwest	10 (25.6%)
Territories	1 (2.6%)

divisions defined by the United States Census Bureau (HHS, n.d.). Due to the small sample size obtained here, it was not plausible to use more narrow regional divisions.

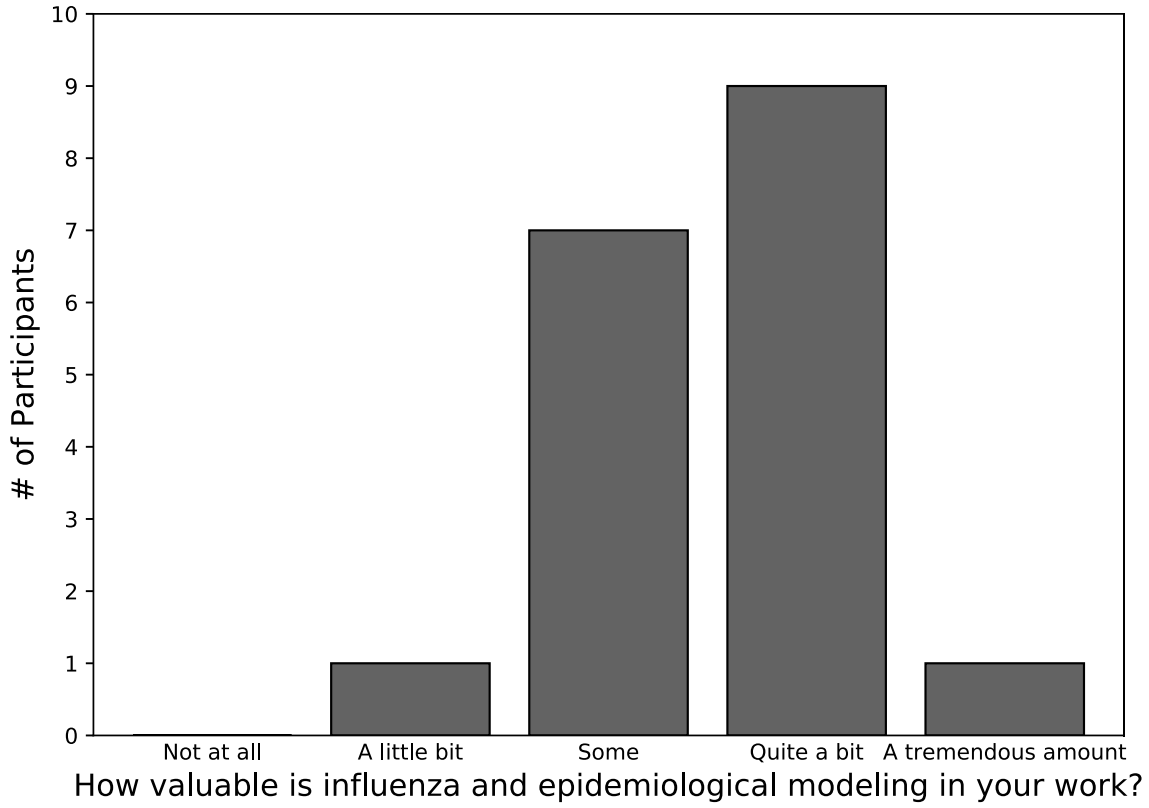


Figure 1. Reported value of models among eighteen public health practitioners who reported using models on the job.

Use of Models

Almost half of respondents (18, 46.2%) reported using models in their work, and that use differed significantly by region (two-tailed Fisher's exact test, P=0.0311; regions are defined as described under "Demographics" above). Specifically, use was highest in the West and lowest in the South and Midwest. Use of models was not significantly related to other demographic variables. Most of these individuals considered the models to be valuable (Figure 1) and used them relatively frequently (Figure 2). Satisfaction with this frequency varied (Figure 2), but was significantly higher with higher frequency of use (two-tailed Fisher's exact test, P=0.003).

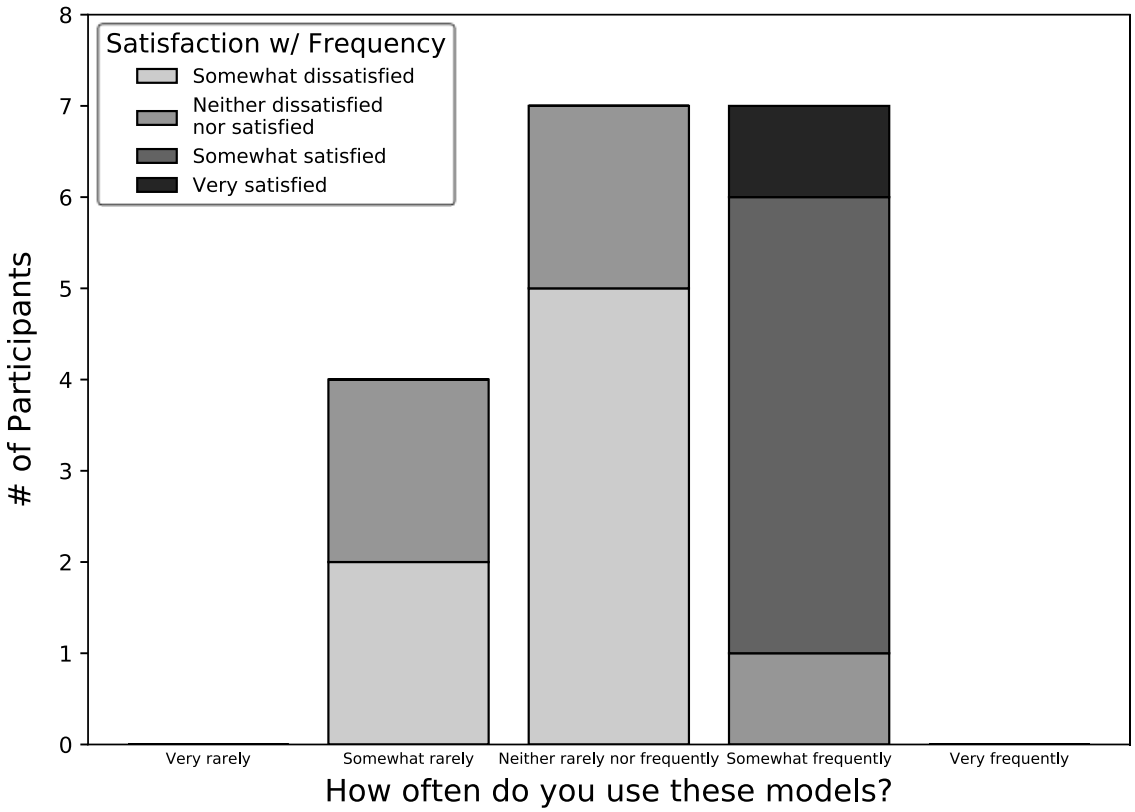


Figure 2. Reported frequency of model use and satisfaction with this frequency among eighteen public health practitioners who reported using models on the job.

Communication with Modelers

A total of eight (20.5%) respondents indicated communication directly with those who develop and create models; seven of these individuals also used models in their own work. Although this interaction occurred fairly rarely (Figure 3), most participants were satisfied with this low level of communication. Again, there was a tendency for satisfaction to be higher with more frequent communication (two-tailed Fisher's exact test, P=0.043), but the sample size (n=8) was very small.

When asked how communication with modelers could be improved, 26 (66.7%) respondents indicated that models should be more relevant to public health questions, 23 (59%) wanted increased frequency of telecommunication, 20 (51.3%) desired more plain language from

modelers, and 13 (33.3%) wanted more face-to-face conversation. Three individuals entered their own responses, which were: “Models designed taking into account US jurisdictions outside the contine[n]tal US,” “Provide more information on the value of models to support questions from other health professionals and the media,” and “Greater availability of models. I did not know these existed.”

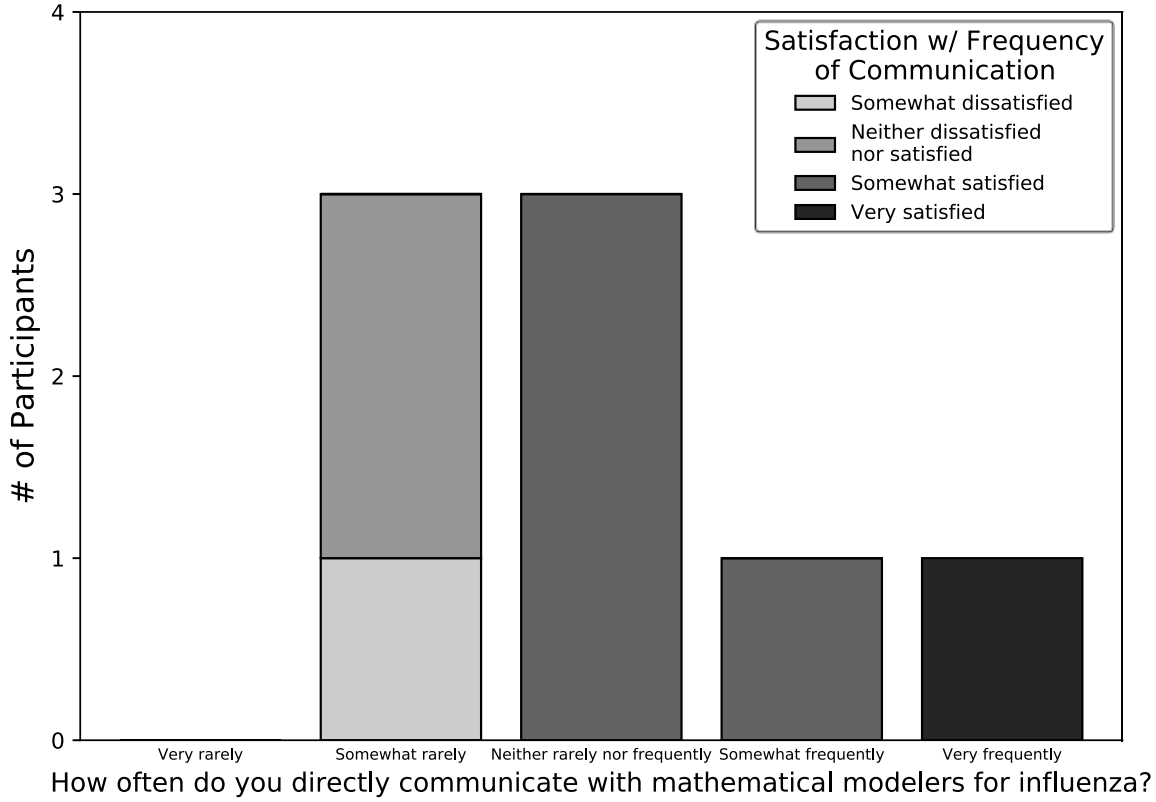


Figure 3. Frequency with which public health practitioners communicated with people who develop mathematical models of influenza and satisfaction with this frequency among eight participants who reported ever communicating.

Awareness of Influenza Forecasts

Twenty-five (64.1%) respondents were aware that forecasts for influenza are available, and 18 (72%) of these individuals had seen one in the past 12 months. These rates were no higher among those who used models in their work than among those who did not (chi-squared test, P=1 and P=0.4423, respectively). Furthermore, participant ratings of model usefulness did

not differ significantly based on whether or not the participant was aware of or had seen forecasts (Fishers exact test, $P=0.509$ and $P=0.597$, respectively.)

Only seven participants (18% of the total; 38.9% of those who had seen a forecast) reported that they or their colleagues had accessed Columbia University's forecasts specifically. Among these seven, three agreed that the forecasts were trustworthy and the other four rated their trustworthiness as neutral. Most (five) said that the forecasts were released neither frequently nor infrequently, and "somewhat frequently" and "very rarely" were also endorsed by one individual each. Finally, only two respondents actually used these forecasts in decision making, with one reporting that the forecasts changed communication strategies with the public and stakeholders and influenced preparedness in a healthcare facility, and the other reporting that the forecasts "supported our regional risk activity assessment."

Data Sources

Of the 39 respondents, 34 (87.2%) reported obtaining influenza incidence or forecast data from the Centers for Disease Control and Prevention (CDC), 14 (35.9%) from Google Flu Trends, five (12.8%) from Columbia University, four (10.3%) from HealthMap FluCast, and seven (18%) from other sources, most commonly state and local ILI (influenza-like illness) reports. Only one respondent reported using no sources at all. However, we note that most (13/14) respondents who said they used Google Flu Trends did so after Google Flu Trends was taken offline in July 2015. Thus, although these individuals used Google Flu Trends in the past, we do not know if they continued to access influenza data from other sources.

Discussion

Despite the potential benefits of using mathematical models, including forecasts, to address public health questions, knowledge of whether and how US-based public health practitioners incorporate model-generated information into decision making is limited. Here, we examine this situation using a cross-sectional survey of 39 public health practitioners in the United States.

Almost half of respondents reported using models in some capacity, and most rated the value of models highly. Future work should determine why some participants view models as more valuable or use models more frequently than others, and efforts should be taken to increase access to, utility of, and user-friendliness of models and model-generated information. Influenza forecasts in particular could be of use to public health practitioners, in that accurate predictions of influenza outbreak metrics, such as peak timing and intensity, could inform vaccination strategies, resource allocation, and communication with the public. Notably, all of our respondents reported that they frequently work with influenza outbreaks. For this reason, it was promising to observe that almost half of surveyed practitioners had seen or had a colleague who had seen an influenza forecast in the past 12 months. However, those who had seen a forecast were not more likely to use models in their work than those who had not, suggesting that these forecasts are not often put to practical use. In fact, although we only asked about forecast use among those who accessed Columbia's forecasts specifically, only two of seven respondents reported actually using the forecasts in public health decision making.

Suboptimal use of available forecasts is an issue in many fields, and is particularly well-studied in agriculture. A study of the use of monsoon forecasts in India found that many farmers complained that forecasts were not available when they were needed, emphasizing the

importance of generating forecasts with appropriate lead time (Martin, 2013). Additionally, a separate review of forecast use in agriculture implicates insufficient forecast quality, both real and perceived, for preventing forecast use in decision making (Kusunose & Mahmood, 2016). Due to the potential severity of influenza, it is logical that the prospect of acting on an inaccurate forecast is concerning to practitioners. Kusunose and Mahmood (2016) suggest that expectations of forecasts might be made more realistic by incorporating the degree of uncertainty associated with predictions, something our group has developed for influenza (Shaman et al., 2013; Shaman & Kandula, 2015). Future studies should further explore the reasons public health practitioners are hesitant to rely on influenza forecasts, as well as the formats and modes of delivery most useful to practitioner work, so that such concerns can be better addressed.

Perhaps our most salient result concerns the overwhelming endorsement of several ways for improving public health practitioner communication with modelers. This finding is in line with previous reviews and qualitative studies (Driedger et al., 2014; Lee et al., 2013; Moghadas et al., 2009, 2015; Moss et al., 2018). To improve communication between modelers and practitioners, both knowledge- and trust-related issues that prevent practitioners from using models effectively should be addressed. For instance, the development of specific guidelines on using mathematical models to answer public health questions may help to clear up misconceptions concerning the capacity of models. Additionally, past qualitative work has found consistency of language and clear communication of model assumptions to be of particular importance (Driedger et al., 2014; Lee et al., 2013; Moghadas et al., 2015). Increased trust in modeling methods and results might also be cultivated by forming longstanding collaborations between practitioners and modelers (Moghadas et al., 2009; Moss et al., 2018). Future work could survey practitioners participating in collaborations with modelers to determine which

communication practices have been most and least effective. While nuanced and detailed communication efforts will be necessary, basic informational campaigns can also play a role: One participant did not know that models existed before taking our survey.

Finally, in addition to the questions posed concerning communication frequency and quality, future surveys should assess how participants communicate with modelers, what topics are discussed, and their endorsement of a variety of ways to improve communication. They should also allow for qualitative responses from participants; these responses could suggest effective methods for increasing communication quality and frequency that may be less obvious to modelers.

Limitations

Despite the novelty of this work, several limitations should be addressed. First, although we attempted to contact a large number of public health practitioners, our response rate was small, making it difficult to draw concrete conclusions, or to statistically assess whether model use differed by variables such as years working in public health. Furthermore, our sample is a convenience sample, and may not be representative of the wider group of public health practitioners. Unfortunately, we know neither the demographic distributions among nonrespondents nor the number of practitioners our survey reached, and can therefore report neither adjusted results nor an overall response rate. However, given that our sample is likely biased toward practitioners with greater knowledge of and interest in mathematical models, we expect that these measures would be even lower among a truly random sample. Thus, our conclusion that model use is below 50%, at least, is likely to hold among US-based public health practitioners in general.

We also note that the definition of the word "model" in our survey was ambiguous. Although we hope that the questions on influenza forecasts prompted participants to think in terms of mechanistic models, it is possible that some respondents took the survey with other types of models, such as regression models, in mind. Similarly, exactly what constitutes model "use" could be anything from simply viewing model output to being actively involved in the development and execution of a model; unfortunately, we cannot tell where each participant falls on this spectrum.

Given that three data points were removed due to inconsistent responses, and that several participants reported using an unavailable data source (GFT), an increased focus on response credibility is indicated. A clear definition of "model" and "model use" will be instrumental in increasing the credibility of future survey results. Reliability can be further enhanced by asking respondents to elaborate their responses, such as through providing a specific categorization or description of the context and form of model used.

Conclusions

Among 39 surveyed public health professionals, both model use and familiarity with influenza forecasts were reported by almost half of participants, but communication with model developers was rare. Improved communication between modelers and practitioners in particular seems to be key for increasing the frequency and effectiveness of model use among public health practitioners. Although more research on why forecasts and other models are not commonly used is necessary, initial improvements should be made now, in the absence of urgent pandemic threats. Participants in a previous qualitative study of eight modelers and public health practitioners noted that effective use of models during the 2009 influenza pandemic suffered

because partnerships between modelers and practitioners were not formed until the pandemic was underway (Driedger et al., 2014). Importantly, communication is a two-way street: Modelers must be more clear about the capabilities and limitations of mathematical models, as well as model interpretation; meanwhile, practitioners must better communicate the information needed from models to better protect the public from outbreaks. Without such communication and use, it is clear that models will not reach their full potential as public health tools.

References

Adde, A., Roucou, P., Mangeas, M., Ardillon, V., Desenclos, J.-C., Rousset, D., Girod, R., Briolant, S., Quenel, P., & Flamand, C. (2016). Predicting Dengue Fever Outbreaks in French Guiana Using Climate Indicators. *PLOS Neglected Tropical Diseases*, *10*(4), e0004681. https://doi.org/10.1371/journal.pntd.0004681

Berg, J. E., Nelson, F. D., & Rietz, T. A. (2008). Prediction market accuracy in the long run. *International Journal of Forecasting*, *24*(2), 285–300. https://doi.org/10.1016/j.ijforecast.2008.03.007

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M. A., Flasche, S., Baguelin, M., Pollington, T., Carney, J. R., Glover, R., Smout, E., Tiffany, A., Edmunds, W. J., & Funk, S. (2015). Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: A Real-time Modelling Study. *PLOS Currents Outbreaks*. https://doi.org/10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2

CDC. (n.d.). *Epidemic Prediction Initiative*. Centers for Disease Control and Prevention. https://predict.cdc.gov/

Chen, I.-F., & Lu, C.-J. (2017). Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications*, *28*(9), 2633–2647. https://doi.org/10.1007/s00521-016-2215-x

Choi, T.-M., Hui, C.-L., Liu, N., Ng, S.-F., & Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, *59*, 84–92. https://doi.org/10.1016/j.dss.2013.10.008

Chowell, G., Hincapie-Palacio, D., Ospina, J., Pell, B., Tariq, A., Dahal, S., Moghadas, S., Smirnova, A., Simonsen, L., & Viboud, C. (2016). Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics. *PLoS Currents*, *8*. https://doi.org/10.1371/currents.outbreaks.f14b2217c902f453d9320a43a35b9583

Columbia University Mailman School of Public Health. (n.d.). *Influenza Observations and Forecast.* Columbia Prediction of Infectious Diseases. https://cpid.iri.columbia.edu/

Debry, E., & Mallet, V. (2014). Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and PM10 on the Prev'Air platform. *Atmospheric Environment*, *91*, 71–84. https://doi.org/10.1016/j.atmosenv.2014.03.049

Driedger, S. M., Cooper, E. J., & Moghadas, S. M. (2014). Developing model-based public health policy through knowledge translation: The need for a 'Communities of Practice.' *Public Health*, *128*(6), 561–567. https://doi.org/10.1016/j.puhe.2013.10.009

FAO. (2016). *Crop yield forecasting: Methodological and institutional aspects.* Food and Agriculture Organization of the United Nations. http://gsars.org/en/crop-yield-forecasting-methodological-and-institutional-aspects/

Gaubert, B., Coman, A., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., & Beekmann, M. (2014). Regional scale ozone data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model. *Geosci. Model Dev.*, *7*(1), 283–302. https://doi.org/10.5194/gmd-7-283-2014

Gneiting, T., & Raftery, A. E. (2005). Atmospheric science. Weather forecasting with ensemble methods. *Science (New York, N.Y.)*, *310*(5746), 248–249. https://doi.org/10.1126/science.1115255

HHS. (n.d.). *Regional Offices*. U.S. Department of Health and Human Services. https://www.hhs.gov/about/agencies/regional-offices/index.html

Hickmann, K. S., Fairchild, G., Priedhorsky, R., Generous, N., Hyman, J. M., Deshpande, A., & Del Valle, S. Y. (2015). Forecasting the 2013–2014 Influenza Season Using Wikipedia. *PLoS Computational Biology*, *11*(5). https://doi.org/10.1371/journal.pcbi.1004239

Huff, A., Allen, T., Whiting, K., Breit, N., & Arnold, B. (2016). FLIRT-ing with Zika: A Web Application to Predict the Movement of Infected Travelers Validated Against the Current Zika Virus Epidemic. *PLOS Currents Outbreaks*. https://doi.org/10.1371/currents.outbreaks.711379ace737b7c04c89765342a9a8c9

Kusunose, Y., & Mahmood, R. (2016). Imperfect forecasts and decision making in agriculture. *Agricultural Systems*, *146*, 103–110. https://doi.org/10.1016/j.agsy.2016.04.006

Lee, B. Y., Haidari, L. A., & Lee, M. S. (2013). Modelling during an emergency: The 2009 H1N1 influenza pandemic. *Clinical Microbiology and Infection*, *19*(11), 1014–1022. https://doi.org/10.1111/1469-0691.12284

Martin, N. M. (2013). *Use of Seasonal Forecast Information in Farm Level Decision Making in Bundelkhand, India*. 11.

Mccarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, *25*(5), 303–324. https://doi.org/10.1002/for.989

Meltzer, M. I., Atkins, C. Y., Santibanez, S., Knust, B., Petersen, B. W., Ervin, E. D., Nichol, S. T., Damon, I. K., Washington, M. L., & Centers for Disease Control and Prevention (CDC). (2014). Estimating the future number of cases in the Ebola epidemic—Liberia and Sierra Leone, 2014-2015. *MMWR Supplements*, *63*(3), 1–14.

Moghadas, S. M., Haworth-Brockman, M., Isfeld-Kiely, H., & Kettner, J. (2015). Improving Public Health Policy through Infection Transmission Modelling: Guidelines for Creating a Community of Practice. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *26*(4), 191–195. https://doi.org/10.1155/2015/274569

Moghadas, S. M., Pizzi, N. J., Wu, J., & Yan, P. (2009). Managing public health crises: The role of models in pandemic preparedness. *Influenza and Other Respiratory Viruses*, *3*(2), 75–79. https://doi.org/10.1111/j.1750-2659.2009.00081.x

Molinari, N.-A. M., Ortega-Sanchez, I. R., Messonnier, M. L., Thompson, W. W., Wortley, P. M., Weintraub, E., & Bridges, C. B. (2007). The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine*, *25*(27), 5086–5096. https://doi.org/10.1016/j.vaccine.2007.03.046

Moss, R., Fielding, J. E., Franklin, L. J., Stephens, N., McVernon, J., Dawson, P., & McCaw, J. M. (2018). Epidemic forecasts as a tool for public health: Interpretation and (re)calibration. *Australian and New Zealand Journal of Public Health*, *42*(1), 69–76. https://doi.org/10.1111/1753-6405.12750

Newlands, N. K., Zamar, D. S., Kouadio, L. A., Zhang, Y., Chipanshi, A., Potgieter, A., Toure, S., & Hill, H. S. J. (2014). An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. *Frontiers in Environmental Science*, *2*. https://doi.org/10.3389/fenvs.2014.00017

Ong, J. B. S., Chen, M. I.-C., Cook, A. R., Lee, H. C., Lee, V. J., Lin, R. T. P., Tambyah, P. A., & Goh, L. G. (2010). Real-Time Epidemic Monitoring and Forecasting of H1N1-2009 Using Influenza-Like Illness from General Practice and Family Doctor Clinics in Singapore. *PLoS ONE*, *5*(4), e10036. https://doi.org/10.1371/journal.pone.0010036

Reich, N. G., Lauer, S. A., Sakrejda, K., Iamsirithaworn, S., Hinjoy, S., Suangtho, P., Suthachana, S., Clapham, H. E., Salje, H., Cummings, D. A. T., & Lessler, J. (2016). Challenges in Real-Time Prediction of Infectious Disease: A Case Study of Dengue in Thailand. *PLoS Neglected Tropical Diseases*, *10*(6), e0004761. https://doi.org/10.1371/journal.pntd.0004761

Shaman, J., & Kandula, S. (2015). Improved Discrimination of Influenza Forecast Accuracy Using Consecutive Predictions. *PLoS Currents*. https://doi.org/10.1371/currents.outbreaks.8a6a3df285af7ca973fab4b22e10911e

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, *109*(50), 20425–20430. https://doi.org/10.1073/pnas.1208772109

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3837

Shaman, J., Yang, W., & Kandula, S. (2014). Inference and forecast of the current west african ebola outbreak in Guinea, sierra leone and liberia. *PLoS Currents*, *6*. https://doi.org/10.1371/currents.outbreaks.3408774290b1a0f2dd7cae877c8b8ff6

Shi, Y., Liu, X., Kok, S.-Y., Rajarethinam, J., Liang, S., Yap, G., Chong, C.-S., Lee, K.-S., Tan, S. S. Y., Chin, C. K. Y., Lo, A., Kong, W., Ng, L. C., & Cook, A. R. (2015). Three-

Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environmental Health Perspectives*, *124*(9). https://doi.org/10.1289/ehp.1509981

Sun, H. (2014). Improving the Effectiveness of Multi-Year Fiscal Planning. *Government Finance Review*, 44–50.

Viboud, C., Boëlle, P.-Y., Carrat, F., Valleron, A.-J., & Flahault, A. (2003). Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, *158*(10), 996–1006.

Yang, W., Cowling, B. J., Lau, E. H. Y., & Shaman, J. (2015). Forecasting Influenza Epidemics in Hong Kong. *PLOS Computational Biology*, *11*(7), e1004383. https://doi.org/10.1371/journal.pcbi.1004383

Zebiak, S. E., & Cane, M. (1987). A model El Niño-Southern Oscillation. *Monthly Weather Review*, *115*, 2262–2278.

Zebiak, S. E., Orlove, B., Muñoz, Á. G., Vaughan, C., Hansen, J., Troy, T., Thomson, M. C., Lustig, A., & Garvin, S. (2015). Investigating El Niño-Southern Oscillation and society relationships. *Wiley Interdisciplinary Reviews: Climate Change*, *6*(1), 17–34. https://doi.org/10.1002/wcc.294

Supplementary Information

Columbia University Pilot Project Survey

Page 1

Thank you for participating in our survey. We are interested in learning more about the use of influenza forecasts in public health decision making. This survey should only take a few minutes to complete, and your answers will help us better understand the use of forecasts. Your responses are confidential. In addition, this survey is voluntary. If you have questions about the survey, please email cfd2113@columbia.edu. Thank you for helping with this study.

Page 2

Influenza forecasts are predictions of future influenza incidence and are generated during the influenza season. They provide predictions of the epidemiological progression of a local outbreak, including local number of weekly cases, outbreak duration, and the week when influenza incidence peaks.

1. Are you aware that influenza forecasts are currently available?

    - Yes
    - No

Page 3

2. Have you or any colleagues seen a flu forecast in the last 12 months?

    - Yes
    - No

Page 4

3. What sources do you use to get influenza incidence or forecast information? Please select all that apply.

    - CDC
    - Google Flu Trends
    - Columbia University
    - HealthMap FluCast
    - I don't use any of these sources listed
    - Other (please specify)

Page 5

4. Have you or your colleagues accessed the influenza forecasts from Columbia University at http://cpid.iri.columbia.edu/flu.html?

    - Yes (sends to Page 6)
    - No (sends to Page 9)

Page 6

5. How have you used the influenza forecasts in decision making at work? Please select all that apply.

    - Reallocated money.
    - Changed communication to public or key stakeholders.
    - Affected preparedness in healthcare facilities.
    - Spurred new research.
    - I have not used them in decision making.
    - Other (please specify)

Page 7

6. The influenza forecasts from Columbia University are trustworthy.

    - Strongly disagree
    - Disagree
    - Neither disagree nor agree
    - Agree
    - Strongly agree

Page 8

7. The influenza forecasts from Columbia University are released

    - Very rarely
    - Somewhat rarely
    - Neither rarely nor frequently
    - Somewhat frequently
    - Very frequently

Page 9

8. Do you use epidemiological models in your work, such as, but not limited to, infectious disease forecasts or quantitative estimates of diseases or conditions?

    - Yes (sends to Page 10)
    - No (sends to Page 12)

Page 10

9. How valuable is influenza and epidemiological modeling in your work?

    - Not at all
    - A little bit
    - Some
    - Quite a bit
    - A tremendous amount

Page 11

10. How often do you use these models?

    - Very rarely
    - Somewhat rarely
    - Neither rarely nor frequently
    - Somewhat frequently
    - Very frequently

11. How satisfied are you with this amount of use?

    - Very dissatisfied

- Somewhat dissatisfied
- Neither dissatisfied nor satisfied
- Somewhat satisfied
- Very satisfied

Page 12

12. Do you directly communicate with those who develop and create the models?

- Yes (send to Page 13)
- No (send to Page 14)

Page 13

13. How often do you directly communicate with mathematical modelers for influenza?

- Very rarely
- Somewhat rarely
- Neither rarely nor frequently
- Somewhat frequently
- Very frequently

14. How satisfied are you with this amount of communication?

- Very dissatisfied
- Somewhat dissatisfied
- Neither dissatisfied nor satisfied
- Somewhat satisfied
- Very satisfied

Page 14

15. Communication between public health professionals and modelers could improve in which ways? Please select all that apply.

- Increased frequency of telecommunication.
- Increased frequency of face-to-face meetings.
- More plain language from modelers.
- Improved relevancy of models to public health questions.
- Other (please specify)

Page 15

16. Do you work in a public health field?

- Yes (send to Page 16)
- No (send to Page 17)

Page 16

17. In which sector do you work?

- Government
- Industry
- NGO
- Academia

Page 17

18. Do your job duties involve planning for, responding to, or dealing with seasonal influenza and its epidemiology?

- Yes
- No

Page 18

19. What is your age?

- 18-29 years old
- 30-49 years old
- 50-64 years old
- 65 years and over

20. What is your gender?

- Female
- Male

21. How many years have you worked in public health and/or epidemiology?

- 0-3 years
- 4-6 years
- 7-10 years
- 10-15 years
- >15 years

22. What is the highest level of school you have completed or the highest degree you have received?

    - Less than high school degree
    - High school degree or equivalent (e.g., GED)
    - Some college but no degree
    - Associate degree
    - Bachelor degree
    - Graduate degree

23. In what state or U.S. territory do you live?

    - Alabama
    - Alaska
    - American Samoa
    - Arizona
    - Arkansas
    - California
    - Colorado
    - Connecticut
    - Delaware
    - District of Columbia (DC)
    - Florida
    - Georgia
    - Guam
    - Hawaii
    - Idaho
    - Illinois
    - Indiana
    - Iowa
    - Kansas
    - Kentucky
    - Louisiana
    - Maine
    - Maryland
    - Massachusetts
    - Michigan
    - Minnesota
    - Mississippi
    - Missouri

- Montana
- Nebraska
- Nevada
- New Hampshire
- New Jersey
- New Mexico
- New York
- North Carolina
- North Dakota
- North Marianas Islands
- Ohio
- Oklahoma
- Oregon
- Pennsylvania
- Puerto Rico
- Rhode Island
- South Carolina
- South Dakota
- Tennessee
- Texas
- Utah
- Vermont
- Virginia
- Virgin Islands
- Washington
- West Virginia
- Wisconsin
- Wyoming

Page 19

24. Have you received the flu vaccine this season?

    - Yes
    - No

25. Have you had influenza in the past five years?

    - Yes
    - No

End

Chapter 6

Conclusion

In this dissertation, we sought to advance the use of forecasting as a tool to reduce morbidity and mortality due to seasonal influenza outbreaks. Specifically, we greatly expanded the geographic range over which skillful influenza forecasts have been generated, and explored the potential for both model improvement and forecast operationalization. In this chapter, we discuss in greater detail the barriers to (1) increasing forecast accuracy and calibration, and (2) practical forecast use. We furthermore make suggestions for future research and collaborations that may help alleviate these barriers. Finally, we briefly discuss implications of this work beyond forecasting, as well as for diseases other than influenza. We organize this discussion into distinct sections accordingly.

6.1: Forecast accuracy and calibration must be improved

Summary of findings (Chapters 2-4)

Chapters 2-4 describe our efforts to develop country-level influenza forecasts using publicly-available data from the World Health Organization (WHO). In Chapter 2, we demonstrated that, although the epidemiologic data from FluID often lacked consistent or meaningful denominator data, accurate and well-calibrated forecasts could still be generated for countries located in temperate regions. Furthermore, we contributed to the evidence base showing that using humidity-forced models significantly improves forecast accuracy in temperate regions (Shaman et al., 2017). Forecasts generated for countries in the tropics and subtropics, however, were significantly less accurate, even when data were smoothed or forecasts were generated for individual outbreaks, indicating the need for improvements in data quality, and for greater understanding of the implications of climatic factors in these regions.

In Chapter 3, we showed that, despite the success of a similar model in the United States (US), introducing connections between countries based on observed commuting and air travel flows was not able to improve forecast accuracy, and in some cases significantly reduced it. We primarily attributed this result to data issues, particularly the lack of denominator data and therefore lack of comparable epidemiologic data between countries. However, we also noted that commuting between countries might not be as important a driver of influenza transmission in Europe as in the US, and furthermore that substantial differences in (sub)type dynamics by country suggest that other factors not captured in our model may play significant roles.

Finally, in Chapter 4, we extended our work from Chapter 2 to generate real-time forecasts of influenza in several temperate countries. During the 2018-19 season the real-time forecasts performed similarly to retrospective forecasts, and on average outperformed methods

based on historical expectance; forecasts for the 2017-18 season were comparable or better than historical expectance. These findings indicate that real-time forecasts could be used to provide national public health organizations valuable information concerning estimated outbreak timing and intensity. In the future, more timely data submissions from reporting countries should be encouraged in order to facilitate production of more real-time forecasts, and continued work is needed to advance forecasting in subsequent seasons.

Improving surveillance data

Model performance is strongly constrained by the quality of available data (George et al., 2019; Keeling, 2005). Throughout Chapters 2-4, we found that data smoothness, and the use of more specific influenza-like illness (ILI) (vs. acute respiratory infection, or ARI) data, was associated with higher forecast accuracy. In Chapter 3, we found that a lack of comparable data between countries limited our ability to build potentially informative metapopulation models (Pei et al., 2018; Yang et al., 2016). In the same chapter, we discussed at length the particular need for denominator data from epidemiologic sentinel systems, so we will not cover this requirement in detail here. We only want to emphasize the importance of collecting visits made to sentinel sites in particular as denominator data. Indeed, the WHO describes such data as “essential” for assessing the burden of influenza (WHO, 2015). While catchment area sizes are more informative than no denominator, they do not account for differences in health seeking behavior, which may be quite substantial between countries. Data on the total number of patients who visit a sentinel site, on the other hand, partially control for these differences.

Barriers to influenza surveillance in the tropics/subtropics

High-quality data are particularly lacking in the tropics and subtropics. In particular, data were either completely unavailable or very low-quality for most countries in Africa. In the Supplemental Materials for Chapter 2, we found that data smoothness was associated with improved forecasts of peak intensity in the tropics, and furthermore that data smoothness was significantly lower in the tropics than in temperate regions. While the regular and systematic collection of ILI and virologic data would undoubtedly improve forecasting capacity in these regions, such a recommendation cannot be made without recognizing the substantial barriers to timely influenza surveillance faced in many countries located in these regions.

In many countries, resources available for public health are scarce, and competing healthcare needs mean that influenza is rarely a priority (Cummings et al., 2016; Katz et al., 2012; Yang et al., 2018). In sub-Saharan Africa in particular, malaria is a prominent competing health priority, and often presents with symptoms similar to those of influenza, making it particularly difficult to recognize influenza cases (Yazdanbakhsh & Kremsner, 2009). This is further compounded by the fact that influenza has only fairly recently been recognized as an important cause of morbidity and mortality in the tropics and subtropics (Viboud, Alonso, et al., 2006; Yazdanbakhsh & Kremsner, 2009). Finally, limited laboratory capacity makes improvements to virologic surveillance difficult (Cummings et al., 2016; Katz et al., 2012; Polansky et al., 2016), while lack of access to healthcare in some regions means that less-severe cases in particular are likely to be missed (Yang et al., 2018).

Given these barriers, and in particular the high prevalence of other infectious diseases such as HIV/AIDS and malaria in some regions (Yang et al., 2018), it seems reasonable that some countries do not currently prioritize high-quality surveillance systems for influenza (Lipsitch et

al., 2011). Nonetheless, researchers emphasize that those countries that can afford to improve their surveillance systems should do so to the best of their ability (Radin et al., 2012; Yazdanbakhsh & Kremsner, 2009). These improvements are not only useful for forecasting, but will also help to better characterize influenza seasonality (or lack thereof) in different regions (Bloom-Feshbach et al., 2013; Ng & Gordon, 2015), improve clinical care among those presenting with nonspecific ILI (Yazdanbakhsh & Kremsner, 2009), and support decision making concerning vaccine choice and timing (Caini et al., 2016; Hirve et al., 2016; Radin et al., 2012).

Potential for successful surveillance improvement

Rigorous surveillance systems collecting high-quality epidemiologic and virologic data require substantial time, money, and personnel investments (WHO, 2011, 2014), and improvement to current systems will not occur in the absence of sufficient political will. However, we do not believe that the recommendations made above and in Chapter 3 are unrealistic. Indeed, substantial improvement over time in the number of countries consistently reporting data to FluNet and FluID is readily observed (Figure 1), although we note that, as the data were downloaded retrospectively, we cannot judge the timeliness of reporting. While much of this improvement was driven by the 2009 influenza pandemic, which led to calls for improved surveillance, particularly epidemiologic surveillance (Ortiz et al., 2009; WHO, 2010), positive trends can be observed already before the pandemic (for FluNet, Figure 1A) as well, indicating that improvement is possible even in the absence of an immediate pandemic threat. In particular, it is encouraging to see substantial increases in reporting over time from countries in the tropics

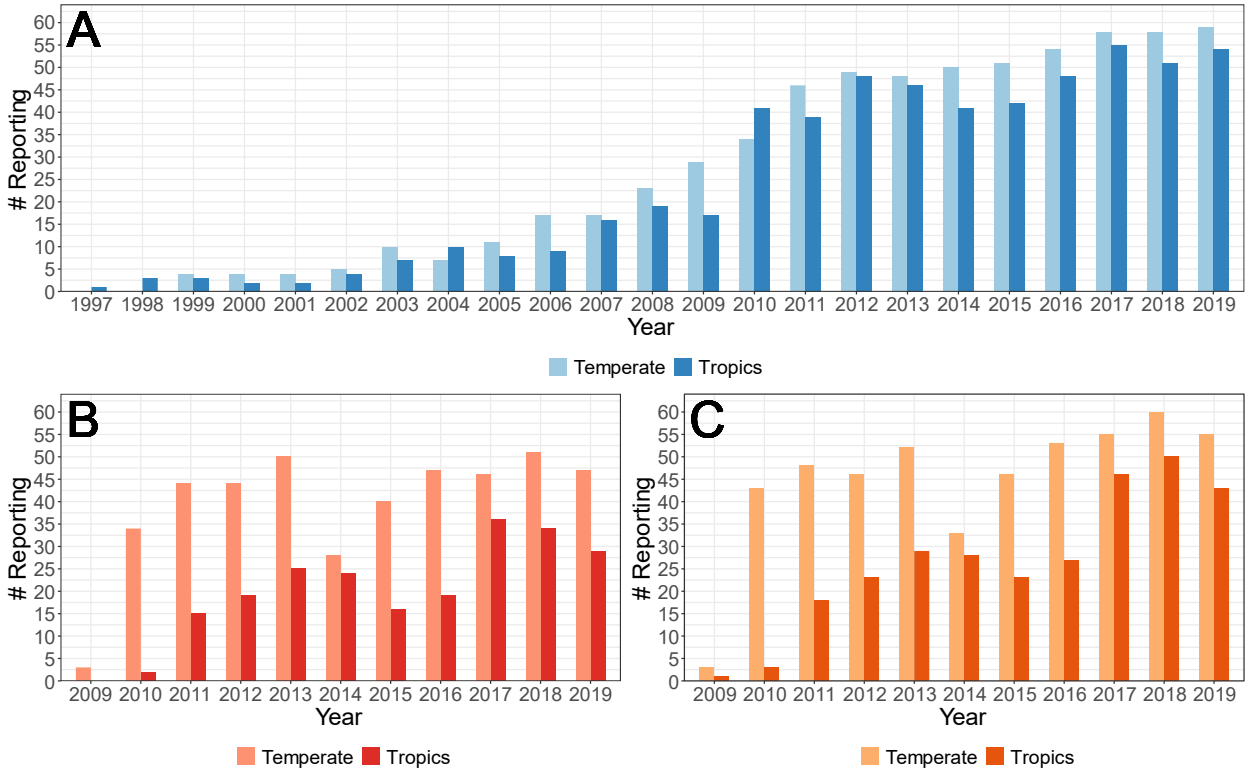


Figure 1. Improvement in reporting of influenza surveillance data to the WHO over time. Bar height shows the number of countries and territories each year that consistently reported (A) positive and total tests for influenza to FluNet; (B) number or rates of ILI to FluID; or (C) number or rates of ILI, ARI, SARI, or pneumonia to FluID. Consistent reporting was defined as submitting data for at least 90% of weeks during the influenza season (weeks 40-19 in the northern hemisphere; weeks 14-45 in the southern hemisphere) in temperate regions, or for at least 90% of week throughout the year in the tropics. Countries and territories were labeled as temperate or tropical as described in Chapter 2. Data through 2015 were downloaded in late 2016; data for 2016-2019 were downloaded in 2020.

and subtropics, despite the barriers discussed above. These improvements were partially driven by partnerships between these countries and the US Centers for Disease Control and Prevention (CDC), the WHO, and other national and international organizations (Polansky et al., 2016; Radin et al., 2012); similar initiatives could be considered to improve surveillance in additional countries. We also note that partnerships between scientists generating forecasts and practitioners using them for decision making could further incentivize the collection of better-quality data, as described more fully below. Finally, we reiterate that we do not know the extent to which data quality in FluNet and FluID are due to issues with collection versus reporting. If some countries

already collect good-quality denominator data, encouraging them to report these to the WHO should be much less challenging than making changes to a country's surveillance system itself.

Improving models of influenza transmission

In addition to better data, forecast accuracy may be improved through the development and use of models that consider additional critical drivers of influenza transmission. In this section, we first cover areas where a better understanding of influenza dynamics and their drivers will be required prior to model improvement, then discuss model improvements that may be made with current knowledge.

Understanding the role of environmental drivers in the tropics/subtropics

By incorporating absolute humidity-forcing into our models, we were able to significantly improve forecast accuracy in temperate countries in Chapter 2; Shaman et al. (2017) previously reached similar conclusions in the US. It is therefore reasonable to assume that a better understanding of the climatic drivers of influenza transmission in the tropics and subtropics could help to close the gap in accuracy observed in Chapter 2. However, a monotonic relationship between absolute humidity and *R0* is insufficient to explain transmission patterns in the tropics and subtropics, where absolute humidity remains high year-round (Tamerius et al., 2011).

A small number of studies have instead suggested that the influence of humidity on influenza survival and transmission is bimodal. Yang et al. (2012), for example, found that influenza viruses survived best at relative humidity values both below 50% and around 100%. Tamerius et al. (2013) found that outbreak peaks in many tropical regions were associated with

"humid-rainy" conditions, suggesting a role for high humidity and precipitation. Finally, Deyle et al. (2016) showed that the negative relationship between absolute humidity and influenza transmission only held below temperatures of about 24°C (75°F); at warmer temperatures, higher absolute humidity appeared to drive higher influenza activity levels.

Notably, the studies (Harper, 1961; Lowen et al., 2007) used by Shaman and Kohn (2009) to demonstrate a monotonic, negative relationship between absolute humidity and the survival and transmission of influenza only explored the impact of humidity either at temperatures below this cutoff, or at 30°C and above, temperatures at which aerosol transmission of influenza has consistently been found to be prohibited (Deyle et al., 2016; Lowen et al., 2007, 2008). For this reason, additional laboratory studies exploring the impact of relative and absolute humidity on virus survival and transmission at temperatures between 20°C and 30°C in particular are needed.

The collection of better-quality influenza and climate (Heaney et al., 2016) data throughout the tropics and subtropics would also facilitate further observational studies comparing observed influenza dynamics and various potential environmental drivers. Finally, modeling studies could assess whether various environmental forcing functions are capable of reproducing influenza patterns in the tropics and subtropics. For example, ongoing work attempts to recreate influenza patterns in Hong Kong using a temperature-mediated absolute humidity forcing function (W. Yang, personal communication). Although such studies cannot confirm the role of different environmental drivers, they can help determine which hypotheses are feasible.

Identifying important drivers of spatial dynamics in Europe

The spatial patterns of influenza outbreaks in Europe are fairly well-characterized. Typically, outbreaks peak earlier in western than in eastern Europe (Paget et al., 2007). As we

described in Chapter 3, circulating subtypes can be similar throughout Europe, but may also differ substantially, even between neighboring countries. The fact that our metapopulation model was unable to improve upon forecasts generated at the individual country level could indicate that our model is misspecified; i.e., it does not properly account for those factors most important in driving these spatial patterns throughout Europe. Here we discuss several potential drivers of influenza transmission throughout Europe and highlight critical gaps in knowledge. While we focus specifically on Europe, we expect that many of the recommendations for further study may also apply more broadly.

Human travel: As described in Chapter 1, the role of various types of human travel on influenza transmission is not well-understood. Unlike in the US, where circulating subtypes are typically similar across regions (CDC, n.d.-b), influenza outbreaks in Europe appear to be less well-mixed between countries, perhaps indicating that between-country travel does not have as strong an influence on transmission patterns as does interstate travel in the US. Thus, while future work should continue to interrogate the role of international travel, including train travel, especially between countries that tend to have similar circulating subtypes, the role of travel on other spatial scales must also be explored.

Given the low rates of cross-border commuting, and the frequency with which dominating subtypes vary by country, we may expect within-country travel to play a dominant role in driving influenza dynamics at the individual country level. Indeed, studies of the spatial patterns of influenza in France have suggested that commuting (Charaudeau et al., 2014), as well as train and auto travel in general (Crepey & Barthelemy, 2007), are associated with influenza synchrony between locations. Similar descriptive studies should be conducted in a variety of countries to

determine how the potential influence of various types of travel differs by country. With the improvements in data collection suggested above, these studies could be performed for both within and between country travel, and the results compared to discover which spatial scale is more important in driving transmission, as well as which types of travel are most important at various spatial scales. Additionally, metapopulation models similar to that described in Chapter 3 could be used to explore how different travel types might influence the spread of influenza throughout a country. Models incorporating commuting, train travel, and automobile travel could be fit to observed influenza outbreaks, and the resulting parameters could be used to run the model in free simulation, iteratively removing different types of travel and observing the impact on outbreak timing and intensity across regions. Of course, such studies would require good-quality subnational data, and will therefore only be possible in countries with geographically and demographically representative surveillance systems, as in France (Flahault et al., 2006; Valleron et al., 1986). The prevalence of such surveillance systems among other European countries is unclear.

Spatial patterns could also be partly dependent on repeated introductions of influenza from outside Europe. Evidence suggests that seasonal outbreaks of H3N2 are primarily caused by strains originating each year in East and Southeast Asia, and not by viruses that persist between seasons (Bedford et al., 2010; Russell et al., 2008), although this is not necessarily true for H1N1 and influenza B (Bedford et al., 2015). It follows that international, and specifically intercontinental, air travel could help drive observed patterns of influenza spread, in that areas that are better connected to non-European countries may be seeded with new influenza strains earlier and more frequently than other locations, potentially leading to earlier outbreaks. Indeed, Brownstein et al. (2006) suggested that international air travel contributes to the timing of

influenza mortality in the US, and Geoghegan et al. (2018) found that outbreak patterns in Australia can be partially explained by multiple introductions from outside the country. Further phylogeographic studies, in which genetic, spatial, and temporal information are considered in an attempt to reconstruct a species' evolutionary history (Lemey et al., 2009), can help improve our understanding of how viruses circulate throughout Europe. In particular, they can demonstrate whether new strains are introduced to Europe and circulate widely, implicating a greater role of travel within Europe, or whether repetitive introductions occur across Europe, implicating air travel on the global scale. In this way, such studies may help inform whether it makes sense to model influenza transmission across Europe as a metapopulation, or whether country-specific models, incorporating seeding from highly-connected countries, might better represent observed influenza dynamics. We note that care must be taken in conducting these studies to avoid contributing to the racism and xenophobia that can be exacerbated by fear of infectious disease introduction from other countries, as observed during the current pandemic of coronavirus disease 2019 (COVID-19) (Haynes, 2020; "Stop the Coronavirus Stigma Now," 2020) and throughout the recent European migrant crisis (Khan et al., 2016).

Climate: As noted above, observed influenza patterns in Europe progress from west to east during most influenza seasons. Due to the geography of Europe, this also implies earlier outbreaks at lower latitudes in many seasons, where humidity is perhaps counterintuitively higher (Figure 2A). Throughout this dissertation, we have treated all European countries as having temperate climates. However, parts of Portugal, Spain, and Italy in particular have more subtropical climates. Thus, while a monotonic humidity forcing function, in which lower humidity leads to higher transmission, is not incompatible with observed influenza dynamics in

Europe, a better understanding of the role of climatic factors in the subtropics, as discussed above, may yield a more complete picture.

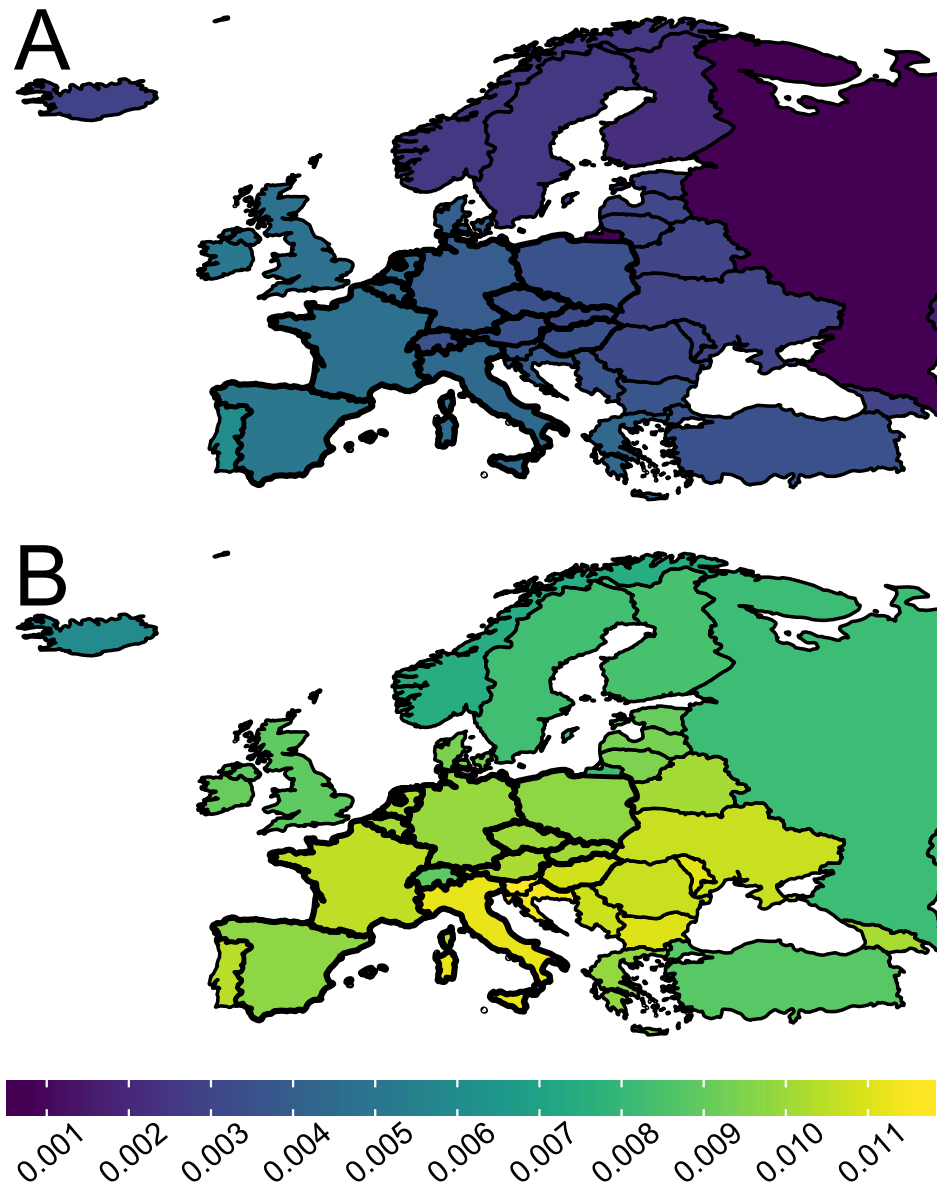


Figure 2. Mean absolute humidity throughout (A) January and (B) July over 20 years in Europe, by country. Data were obtained from NASA's Global Land Data Assimilation System (GLDAS), and aggregated to the country level as described in Chapter 2. The twelve countries included in the metapopulation model in Chapter 3 are outlined in bold. Humidity values are shown in kg/kg.

Here, the cases of Brazil and Australia, two countries comparable in size to Europe and spanning wide geographic ranges, may be illustrative. Alonso et al. (2007) observed that seasonal outbreaks of influenza in Brazil had earlier peaks in the northern, subtropical states than in the southern, temperate states. This pattern was observed despite higher population density

and rates of human travel in the south, suggesting that climatic and not demographic drivers were responsible. Future studies on the role of climate in the subtropics, as discussed above, may help reveal if similar mechanisms are responsible in Brazil and Europe. In Australia, substantial between-season circulation was found to occur in subtropical and tropical regions (Geoghegan et al., 2018), occasionally even seeding new seasonal outbreaks (Patterson Ross et al., 2015). Currently, influenza surveillance during the summer is absent or greatly reduced in most European countries. Increased surveillance would allow for the study of out-of-season circulation patterns, which would reveal whether positivity rates are higher in subtropical regions of Europe during the summer, as well as whether subtropical regions seed seasonal outbreaks, or are simply able to sustain transmission of seeded strains earlier than temperate regions. Based on patterns of summer humidity, regions on the Adriatic Sea may be of particular interest (Figure 2B), although subtropical climates in Spain and Portugal should also be explored. While persistence of H3N2 influenza between seasons is rare (Bedford et al., 2010; Russell et al., 2008), continual circulation of H1N1 and B influenza between seasons (Bedford et al., 2015; Patterson Ross et al., 2015) could allow influenza in these regions to seed outbreaks in temperate Europe.

Population susceptibility: As we have observed, (sub)type circulation patterns often differ between European countries. Historical patterns of (sub)type circulation, coupled with differences in outbreak intensity and vaccine uptake, could lead to complex and highly varying immunity profiles in different countries. In both Chapters 2 and 3, our model displayed a tendency to infer higher values of initial population susceptibility in countries in southwest Europe, although whether this reflects reality or was simply the model's attempt to fit observed geographic patterns by assigning fewer susceptible people to locations with lower absolute

humidity is unclear. Serological surveillance, a pilot of which was recently proposed in the United Kingdom (UK) (de Lusignan et al., 2019), could help to reveal the extent to which overall and (sub)type-specific immunity differs by country. Recording of influenza vaccine uptake among healthy adults (rather than only among high-risk groups (ECDC, 2018)) would also be informative.

Sociodemographic factors: Finally, a substantial body of work has identified several sociodemographic factors that may play a role in driving influenza incidence patterns. Age in particular may be an important driver. Contact patterns are highly assortative by age (Mossong et al., 2008), and children are especially likely to contribute to transmission (Lau et al., 2015; Viboud, Boëlle, et al., 2004). Indeed, past modeling studies have suggested that school holidays can help explain spatial patterns of both seasonal and pandemic influenza outbreaks. Specifically, simultaneous school vacations can drive increased outbreak synchrony between locations (Ewing et al., 2016), while school vacation timing in conjunction with differences in specific humidity can exacerbate outbreak timing differences between locations (Tamerius et al., 2015). Meanwhile, analyses of age patterns among local and imported cases during the 2009 pandemic suggested that adults may be most likely to transmit influenza between locations (Apolloni et al., 2013), a conclusion also supported by studies showing a strong influence of commuting on transmission patterns (Pei et al., 2018; Viboud, Bjørnstad, et al., 2006). For these reasons, differences in age structure between populations could contribute to observed influenza transmission patterns. Notably, FluNet and FluID report age-specific data, which may allow models to estimate rates of transmission within and between age groups in different countries. Detailed data on contact rates by age could help to validate results (Mossong et al., 2008).

However, we have not explored the extent to which good-quality, age-specific data are reported by different countries, and the lack of denominator data may hinder comparability between age groups.

Differences between countries in the distribution of other sociodemographic factors could also help to explain differences in outbreak timing and intensity. For example, city population size has been previously associated with both the extent to which influenza outbreaks are concentrated in time (Dalziel et al., 2018) and with synchrony between locations (Morris et al., 2018). Thus, outbreak patterns could depend partially on the number, size, and population density of urban population centers. Socioeconomic factors, including educational attainment and poverty rates, also vary substantially throughout Europe (European Commission & Statistical Office of the European Union, 2018). Although research on the relationship between socioeconomic status (SES) and influenza is sparse, particularly in Europe, a study of hospitalization rates in Spain during the 2009 pandemic found that risk of hospitalization was lower among those with secondary or higher education (Mayoral et al., 2013), and several studies have reported a positive association between influenza vaccination and both education and income (Lucyk et al., 2019). Future studies must continue to assess the extent to which geographic differences in various sociodemographic factors drive differences in influenza outbreak characteristics; results of such studies have important implications not only for modeling, but for public health in general (Semenza, 2010).

Improving model specification using existing knowledge

While a better understanding of the points discussed above is important for eventually constructing better models, it is important to also consider what modelers can do now to improve model specification.

(Sub)type-specific forecasts: In Chapter 3, we describe (sub)type-specific forecasts motivated in part by observed differences in dominant (sub)types between countries. However, (sub)type-specific forecasts may also improve forecast quality at the isolated country level. The SIRS model describes the transmission dynamics of a single pathogen; thus, modeling individual influenza (sub)types, rather than total influenza, improves model specification. Previous work has shown that, when forecasting total influenza in US cities and states, the sum of (sub)type-specific forecasts generally outperformed forecasts treating influenza as a single pathogen (Kandula, Yang, et al., 2017). Moving forward, similar methods should be tested at the country level. Distinguishing between (sub)types may be particularly relevant in the tropics and subtropics, where lack of cross-immunity between (sub)types may help account for sequential outbreaks of influenza.

Incorporating school schedules: As discussed above under "Sociodemographic factors," the incorporation of school schedules into models of influenza transmission can improve model agreement with observed influenza activity (Ewing et al., 2016; Tamerius et al., 2015; Yaari et al., 2016). Thus, it makes sense for future work to assess the impact of accounting for school holidays in the models described in this dissertation, at least in countries where the timing of school holidays is consistent across regions. School holidays could be implemented simply by

reducing β multiplicatively during the corresponding calendar weeks; the exact extent of this reduction could be fit. If age structure is incorporated into future models, a reduction in β between children specifically would become possible (Ewing et al., 2016; Yaari et al., 2016).

Explicit observation models: The wide range of case counts observed in the WHO data suggest that surveillance systems and health seeking behavior vary extensively by country. Models that explicitly distinguish between influenza cases that are observed by the healthcare system, and those that remain unobserved, can more directly account for these differences than the wide range of scaling factors we have employed throughout this work. At the country level, such models can be constructed by simply dividing the infected compartment (I) into observed and unobserved infecteds. The percentage of cases who are observed would depend on parameters describing the size of the population under surveillance and the rate of healthcare seeking for influenza among this population (Yaari et al., 2016). Priors for the latter parameter could be based on published rates at which individuals infected with influenza seek medical attention (Galanti et al., 2020; Metzger et al., 2004), recognizing that both of the cited studies were limited to New York City. The size of catchment areas is freely available in some countries (WHO & ECDC, 2019) but in many cases may require contacting public health workers from each country.

Accounting explicitly for case observation rates may not only improve model specification (Osthus et al., 2019), but the inferred values of rates of healthcare seeking in different countries could be compared to help understand how such behaviors may vary by country, as well as season to season (Moss, Zarebski, Carlson, et al., 2019). This will, however, depend on how well the models are able to fit this parameter. As observed in our sensitivity

analyses in Chapter 3, as well as in the literature (Pei et al., 2018; Yang et al., 2014), some model parameters are easier to accurately estimate than others. Synthetic testing, both for isolated countries and the metapopulation model, will help to determine whether or not this is likely to be a plausible approach.

Alternative approaches to calculating scaling factors for a network model: Should a metapopulation model explicitly incorporating observed vs. unobserved influenza cases prove implausible, alternative approaches to implicitly account for differences in surveillance systems and rates of healthcare seeking may also be explored. Pei et al. (2018), for example, calculated scaling factors based on the intensity of synthetic outbreaks generated using their metapopulation model in each state. A similar approach could be attempted using the metapopulation model in Chapter 3. Obtaining more information on the size of catchment areas in each country, as discussed above, could also inform the choice of scaling factors.

Exploring alternative filters and models

Throughout this dissertation, all fitting and forecasting work was performed using a mechanistic SIRS model in conjunction with an EAKF. However, countless other methods exist for producing influenza forecasts. Yang et al. (2014), for example, assessed the capabilities of six different filtering methods, including the EAKF, in fitting and forecasting influenza in several US cities, and found that, in general, particle filters were superior to ensemble filters at forecasting peak timing prior to the predicted peak. While particle filters are not a plausible choice for the metapopulation model in Chapter 3 due to the high dimensionality of the model

(Snyder et al., 2008), these methods could be tested in isolated, country-level forecasts, and results compared to those found in Chapters 2 and 4.

In future work, the potential of alternative model forms (e.g. the SEIRS model, which explicitly accounts for individuals who are exposed but not yet infectious) could also be explored (Yang et al., 2014). Furthermore, various statistical methods are capable of generating good-quality influenza forecasts. These methods are non-mechanistic, meaning that they do not explicitly model the transmission of influenza throughout a population. The method of analogues, which we have used and discussed briefly in Chapters 2 and 4, is one such method, and has been shown to be capable of producing accurate influenza forecasts in France (Viboud et al., 2003). Other statistical methods employing time series models, regression analyses, Bayesian approaches, machine learning methods, and combinations thereof have also been used to generate influenza forecasts (Brooks et al., 2015; Kandula et al., 2018; Soebiyanto et al., 2010; Su et al., 2019; Wang et al., 2015). To our knowledge, much of the WHO data used throughout this dissertation have not been used in forecasting efforts outside of this work; for this reason, it would be useful for future studies to compare a range of forecasting methods, both mechanistic and statistical. Given that the method of analogues performed similarly to or better than our model-inference system for tropical countries in Chapter 2, non-mechanistic methods may be of particular utility in these regions. Methods like the method of analogues or the Bayesian weighted outbreaks method, which base predictions on patterns observed in previous outbreaks, are also likely to produce better forecasts of outbreak onset, at least in isolated models (Kandula et al., 2018).

Finally, while comparing the potential of the various forecasting methods discussed above, both overall and by country, is certainly important, optimal forecast performance is likely

to be achieved when multiple methods are considered simultaneously. Superensembles combine forecasts from several models by weighting according to forecast performance in a training dataset (Yamana et al., 2017). Recent work has shown that, in the US, superensembles perform similarly to or better than a wide range of individual forecasting methods; in particular, forecast quality tends to be more consistent across lead weeks (Kandula et al., 2018; Ray & Reich, 2018; Yamana et al., 2017). Future work should assess the performance of superensemble forecasts of influenza in different countries. Furthermore, comparing the weights assigned to different component models between countries and regions could be helpful in understanding geographical differences in outbreak dynamics.

Incorporating error correction and post-processing

In a nonlinear system such as the SIRS model, rapid error growth can substantially degrade forecast accuracy, even at short timescales (Pei & Shaman, 2017). The use of data assimilation methods alone does not fully account for the resulting error. A handful of recent studies have shown that, by incorporating methods of error correction commonly used in weather and climate forecasting, accuracy and calibration of influenza forecasts can be further improved. More specifically, Pei and Shaman (2017) used error breeding methods in conjunction with an EAKF to improve forecast accuracy. Later, Pei et al. (2019) showed that widening ensembles using optimal perturbation of ensemble members improved both the accuracy and calibration of short-term forecasts, particularly around the time of the predicted peak.

Error impacting forecast accuracy may also stem from model misspecification. Osthus et al. (Osthus et al., 2019) developed a model accounting for historical discrepancies between observed data and best-fit SIR models, which produced high-quality forecasts of ILI in the US.

Finally, forecasts may additionally benefit from ad-hoc adjustments made based on modeler and practitioner expertise (Kandula et al., 2018; Morita et al., 2018). For example, weights allocated to different values for peak timing, peak intensity, and onset timing may be edited by modelers after forecasts are produced. This may be done if, for example, experts believe that the forecast has placed too much confidence in a given bin, or to avoid bins receiving a weight of zero (Kandula et al., 2018).

Future work should explore the utility of various error correction and post-processing methods for improving the forecasts discussed in this dissertation. The potential of these methods to improve not only accuracy but also calibration may be of particular interest in the case of the metapopulation model, which on average produced forecasts of peak timing with lower MAE than the isolated models, but also with lower log scores, indicating calibration issues.

Determining how spatial scale impacts forecast accuracy

Throughout this dissertation, we have produced forecasts at the country level. As noted in Chapter 2, these forecasts were generally comparable in quality to forecasts made for individual cities in the US (Shaman et al., 2013) and Australia (Moss et al., 2017). Furthermore, absolute humidity aggregated to the country level was still able to improve forecast accuracy. Finally, we did not observe any noticeable patterns in forecast accuracy by country size. Altogether, these results suggest that our forecasts did not suffer due to the larger spatial scale at which they were generated.

Nevertheless, several researchers have expressed a need for good-quality forecasts at smaller spatial scales (Kandula, Hsu, et al., 2017; Morita et al., 2018; Muscatello et al., 2017), which may be more practical and relevant for mounting a public health response. For this reason,

future work should explore how spatial scale impacts forecast quality in countries where data on several scales are available, and how differences in data quality at different scales contribute to differences in forecast quality. These questions are relevant not only to isolated forecasts, but also to forecasts using improved network models: a metapopulation model incorporating commuting flows was found to improve forecast accuracy in New York City when influenza was modeled at the borough scale, but often reduced accuracy when modeled at the smaller, neighborhood scale (Yang et al., 2016). Results of such studies may help motivate higher quality and more widespread surveillance at local scales and could inform decisions on the ideal spatial scale of real-time forecasts. Of course, such decisions will also be informed by collaborations with public health and medical practitioners, as discussed below, as practitioners will have insight into the spatial scales at which forecasts are likely to be useful.

Exploring the potential of nowcasting methods

Although the real-time forecasts described in Chapter 4 were relatively accurate and well-calibrated, data are submitted to the FluNet and FluID databases with a one-week lag, meaning that forecasts are not truly generated in real-time. While more timely submission of traditional surveillance data may be impractical, various alternative data sources, including Google search data (Kandula, Hsu, et al., 2017; Lampos et al., 2015; Lu et al., 2019), Wikipedia page visits (Zimmer et al., 2018), and tweets (Paul et al., 2014), are published in near real-time and can be used to produce nowcasts, or estimates of current influenza incidence (Lampos et al., 2015). Using such nowcasts to estimate influenza incidence during the week at which a forecast is generated can significantly improve real-time forecast accuracy (Kandula et al., 2019; Paul et al., 2014; Zimmer et al., 2018). However, nowcasting models have primarily been validated against

influenza surveillance data from the US. While there has been some work exploring the association between Google search and influenza surveillance data in several European countries (Samaras et al., 2017; Valdivia et al., 2010; Vandendijck et al., 2013), further research to adapt nowcasting methods to these countries is needed. Specifically, future work should explore whether there are differences in the predictability of Internet search behavior by country, and to what extent the use of nowcasts improves real-time forecasts of influenza. We note, however, that results may be tentative at this point, as we only have access to two seasons of real-time data.

6.2: Options for operationalizing forecasts must be pursued

Summary of findings (Chapter 5)

In Chapter 5, we describe the results of a small, cross-sectional survey of US public health practitioners working with influenza. We found that, while many respondents were familiar with models and forecasts, both communication with modelers and use of forecasts to support decision making were uncommon. Most respondents indicated that increased communication between modelers and practitioners was needed, but also that modelers needed to communicate more clearly, conclusions that agree with similar past studies (Driedger et al., 2014; Lee et al., 2013; Moghadas et al., 2015). Respondents also indicated that models needed to be more relevant to public health. We concluded by emphasizing the importance of longstanding collaboration between modelers and public health practitioners in achieving these goals, and in effectively operationalizing forecasts.

Meaningful collaborations with public health practitioners and clinicians must be built and maintained

As discussed in Chapter 5, several others have emphasized the need for longstanding, formal collaborations between modelers and public health practitioners (Driedger et al., 2014; Moghadas et al., 2015; Rivers et al., 2019). Here, we suggest that clinicians should also be involved in such collaborations. Figure 3A highlights the main contributions that modelers, practitioners, and clinicians can make to these partnerships, as well as the benefits gained by each group. We describe how such a community would function in more detail here.

In such collaborations, one of the roles of public health practitioners (including epidemiologists, data analysts, and policymakers (Lofgren et al., 2014)) and of clinicians (including department heads and hospital administrators) is to determine which forecasting targets are most relevant to their ability to respond to an outbreak (Driedger et al., 2014; Lutz et al., 2019; Metcalf et al., 2015; Moghadas et al., 2015). We and others (Biggerstaff et al., 2018, 2019; Farrow et al., 2017; George et al., 2019; Moss et al., 2018; Shaman et al., 2013; WHO, 2017) have emphasized the potential that skillful forecasts have to guide responses taken by both public health practitioners and medical professionals during an influenza outbreak. By involving these stakeholders in planning for forecast outcomes, this potential is more likely to be met.

Practitioners and clinicians can also offer important insights into the data used for forecasting (Biggerstaff et al., 2019; Moss et al., 2018). While some information on surveillance systems is publicly available (WHO & ECDC, 2019), the exact process behind data collection is not obvious. The insider knowledge provided by public health practitioners and especially clinicians, who collect the primary epidemiologic and virologic data, can reveal attributes and biases of the data that would otherwise have remained unclear. This knowledge can be used to

A

	Public Health/Medical Professionals	Modelers
Contributions	• Communication of which forecast outputs are most relevant to public health practice • Collection of data • Insight into data, including biases	• Communication of model capabilities, limitations, and assumptions • Modeling expertise • Insight into model results, including certainty
Benefits	• Forecasts with higher applicability • Better understanding of forecast results • Better ability to apply forecasting results in decision making	• Better access to data in real-time • Better understanding of data used in forecasting • Increased trust from practitioners

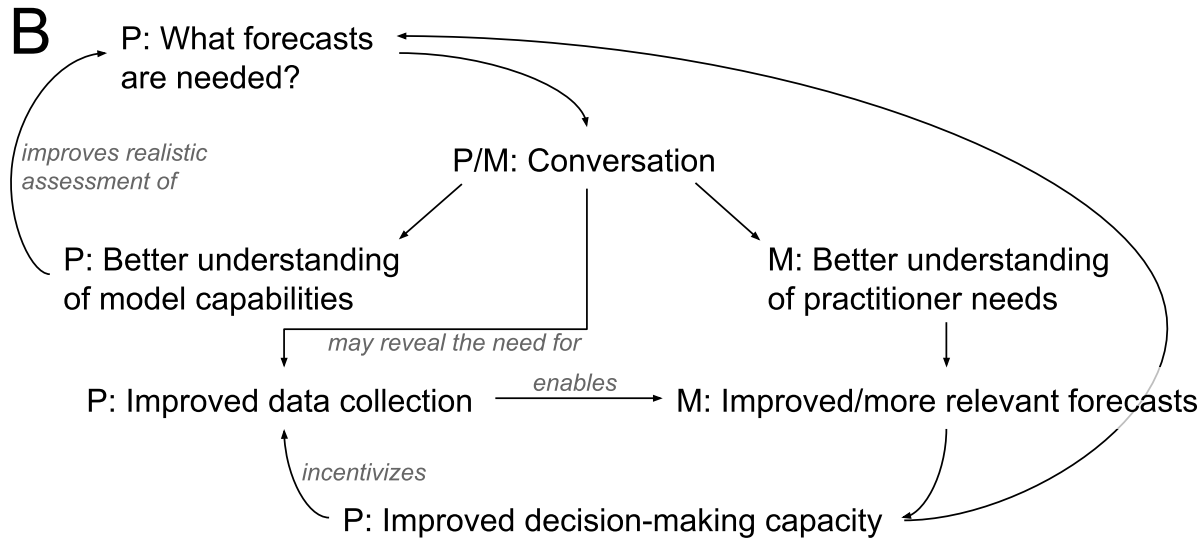


Figure 3. Summary of collaborations between modelers and public health and medical practitioners. (A) Contributions of and benefits to public health practitioners and medical professionals (left) and modelers (right). (B) Potential positive feedback loops resulting from longstanding collaboration.

improve forecast accuracy, as observed in a collaboration between Moss et al. (2018) and Australian health departments, where practitioner knowledge of increases in influenza testing allowed the authors to edit model parameters accordingly mid-season. Input from both clinicians and public health practitioners is also likely to be particularly critical when attempting to consolidate data from different regions or countries, as in Chapter 3. Even if data on outpatient visits were available, differences in sentinel surveillance systems in different countries could still

lead to noticeable biases in resulting syndromic+ rates. A better understanding of these differences may enable a more appropriate accounting for these biases.

Meanwhile, modelers involved in these collaborations have the responsibility to clearly communicate not only the general abilities and limitations of forecasting models (Metcalf et al., 2015), but also the parameter requirements and assumptions of the specific models being used (Muscatello et al., 2017). Modelers will also need to explain and contextualize the results from any forecasts generated (Lutz et al., 2019; Muscatello et al., 2017). (We discuss the communication of model certainty in more detail below.) These contributions will enable meaningful interpretation and utilization of forecast results by non-modelers (Moghadas et al., 2015). We note that even relatively simple forecasting models can require modeling expertise to properly parameterize (Morita et al., 2018). Thus, it is not enough for modelers to simply create models, explain their use, and disseminate them to public health agencies. Rather, modelers must be involved in the actual generation of the forecasts themselves. This involvement has the added benefit of allowing model adjustments to be made mid-outbreak, as done by Moss et al. (2018).

Importantly, the above contributions alone are insufficient; participants will need to engage in significant back-and-forth conversation to truly achieve the goals of the collaboration (Driedger et al., 2014; Moghadas et al., 2015). For example, modelers will need to assess whether the goals put forward by practitioners and clinicians are practical (Driedger et al., 2014). If they are not, modelers will need to explain why, and assist in editing the goals. These discussions may also motivate additional data collection, if certain data are required to produce forecasts capable of influencing policy (Lofgren et al., 2014; Moghadas et al., 2015; Van Kerkhove & Ferguson, 2012). In this way, initial ideas can be carefully refined based on both

practitioners' experience concerning what information is useful in responding to seasonal outbreaks, and modelers' knowledge of the capabilities of various models used for forecasting.

While such discussions may be difficult at first, over successive influenza seasons, modelers will gain a better understanding of the needs of practitioners and clinicians, and practitioners and clinicians will gain more insight into what models can and cannot do. This will not only contribute to increased trust in models among non-modelers, but also has the potential to initiate several positive feedback loops capable of further advancing both the quality of forecasts and their use in practice (Figure 3B). As practitioners gain a greater understanding of models and their capabilities, they will be better able to identify public health questions to which forecasts are applicable. Similarly, the expansion of forecasts with direct relevance to public health, and the resulting improvement in decision-making ability, can encourage the generation of novel policy questions, leading to novel forecasting innovation, and so on (Lofgren et al., 2014; Metcalf et al., 2015). These experiences could be particularly informative in more strategically utilizing forecasts during future pandemics (Driedger et al., 2014; Lee et al., 2013), perhaps avoiding some of the mistakes made during the current COVID-19 pandemic (see "Precedent for using forecasts in decision making" below). Finally, developing forecasts based on extensive stakeholder input would enable demonstration of the potential for forecasts to produce relevant, actionable results, and would help legitimize forecasts in the eyes of public health practitioners and clinicians (Driedger et al., 2014). This in turn has the potential to motivate the collection of better-quality surveillance data (Biggerstaff et al., 2019), which itself allows for better-quality forecasts to be produced. Here again the inclusion of clinicians is of particular relevance: clinicians are responsible for collecting surveillance data, but may not see any direct benefits from their efforts, limiting the incentive to spend additional time and

resources improving surveillance. Forecasts that are of use to clinicians have the power to change this (ECDC, 2014).

While we do not know of any longstanding communities matching what we have described here, several nascent communities and collaborations exist involving both public health practitioners and forecasters. In Chapter 1, we mentioned the Influenza Incidence Analytics Groups (IIAG), a community of modelers and practitioners (including the author of this dissertation) organized by the WHO with the broad goal of advancing the use of WHO influenza data in real-time (Biggerstaff et al., 2019). We discuss above the ongoing collaboration of modelers and health departments in Australia, and the ways in which these communities have led to improvements in forecast skill (Moss et al., 2018; Moss, Zarebski, Dawson, et al., 2019). Other communities include a collaboration between the Council of State and Territorial Epidemiologists (CSTE) and the CDC's Epidemic Prevention Initiative (EPI) (CDC, n.d.-a; *The Epidemic Prediction Initiative*, n.d.) beginning in 2017, which is exploring ways of using forecasts in public health decision making (Lutz et al., 2019); and mod4PH (*Mod4PH*, n.d.), an online network of modelers and public health practitioners dedicated to discussing the potential of mathematical models in public health decision making (Milwid et al., 2016; *Mod4PH*, n.d.).

Improving the calculation and communication of forecast certainty

We have already discussed the importance of forecasts being not only accurate, but well-calibrated. However, it is also vital that we communicate information on certainty clearly to stakeholders. Although their work focused on weather and not infectious disease forecasts, Kusunose and Mahmood (2016) suggest that providing meaningful and clear information on certainty is critical in driving practical forecast use, and Van Kerkhove and Ferguson (2012)

observed that unclear communication of uncertainty hindered the practical use of models during the 2009 influenza pandemic.

Despite this importance, it is unclear exactly how uncertainty can best be communicated to decision makers. Indeed, different researchers have chosen a variety of certainty metrics to present (Columbia University Mailman School of Public Health, n.d.; Moss et al., 2018; Shaman & Kandula, 2015). It is important that certainty is presented in such a way that practitioners can make informed and contextualized decisions, but without including so much information that the big picture becomes obscured. Crucially, forecast calibration also plays a role. As discussed in Chapters 2 and 4, a 90% prediction interval based on ensemble spread, for example, may not contain exactly 90% of observed values, and is therefore not a perfectly honest indication of certainty. A method similar to that reported by Shaman and Kandula (2015), in which forecast calibration is assessed over several seasons to determine which prediction intervals actually capture different percentages of observed values, will therefore be important for communicating certainty accurately. These findings would then need to be updated each year as more data become available.

Insight into how best to communicate forecast certainty will naturally emerge from collaborations of the sort described above, but even in the absence of such communities, future work should explore how different ways of communicating certainty are received by practitioners. An ongoing CDC project exploring different ways to visualize forecasting results may provide initial information (Lutz et al., 2019).

Precedent for using forecasts in decision making

As infectious disease forecasting is a relatively new field, there are not currently many examples of forecasts being integrated into decision making (Morgan, 2019; Rivers et al., 2019), although this is beginning to change (George et al., 2019). As discussed previously, Moss et al. (2018; 2019) collaborate with state- and city-level public health departments not only to share forecasting results but also to gain valuable feedback on model assumptions and communication of results. Furthermore, forecasts generated as part of the CDC's FluSight initiative, a component of EPI, are used by the CDC to broadly evaluate risk in the upcoming weeks, which is then used in both internal and public communication (CDC, n.d.-a; Lutz et al., 2019; *The Epidemic Prediction Initiative*, n.d.). In this way, forecasts are beginning to inform practitioners' and the public's expectations about how outbreaks will unfold. However, this is the extent to which real-time forecasts are used during seasonal outbreaks of influenza at present. Thus, it is helpful to briefly consider the use of real-time forecasting for other infections, namely Ebola and COVID-19.

Early real-time projections of the 2014-2016 Ebola epidemic, showing that the number of cases would increase dramatically in the absence of proper control strategies (Meltzer et al., 2014), played an important role in urging governments to take action (Funk et al., 2019; Morgan, 2019; Rivers et al., 2019). (Here we describe simulations as "projections" rather than "forecasts," as resulting predictions are based on considerable assumptions regarding which public health measures will or will not be taken.) Although early projections had a tendency to overestimate peak intensity, this can be at least partly explained by ensuing actions to control outbreak spread (Chowell et al., 2017); in other words, we have no way of knowing how accurate the predictions would have been had no mitigating actions been taken. As incidence decreased, real-time

forecasts showed that the number of expected future cases would likely be too low to support randomized controlled trials of potential Ebola vaccines, and thus contributed to the development and use of novel vaccine study designs (George et al., 2019; Mohammadi, 2015; "The Ring Vaccination Trial," 2015). While these successes are encouraging, the lack of prior communication and collaboration between modelers and decision makers made the practical use of forecasts difficult (Rivers et al., 2019). In particular, low data quality presented a significant obstacle to the development of skillful real-time forecasts (Chowell et al., 2017).

Many groups and individuals are developing real-time projections for the current COVID-19 pandemic, and it is difficult to assess the quality and policy impact of each one. However, broadly speaking, projections generated in collaboration with national governments have had substantial policy impact, contributing to decisions to implement extensive school and workplace closures, border control, and other social distancing measures (Adam, 2020; Enserink & Kupferschmidt, 2020). Thus, these projections have contributed to policy decisions that have drastically reduced morbidity and mortality. However, early projections of COVID-19 in the UK initially contributed to the adoption of less-severe control measures, before updates to model parameters revealed that these measures would be insufficient, leading to the uptake of additional measures (Adam, 2020). Some uncertainty in projection results is to be expected during an emerging pandemic, when information needed to parameterize models is lacking. However, this means that caution must be taken when applying results to policy, as decisions based on inaccurate projections not only risk contributing to increased morbidity and mortality, but also degrade trust among both public health practitioners and the general public. Indeed, there has already been substantial criticism of what is perceived as the UK's overreliance on projections in making these decisions (Alwan et al., 2020; Enserink & Kupferschmidt, 2020). Thus, while it is

encouraging that projections have seen practical use in reducing the burden of COVID-19 worldwide, whether public perception of projections and of forecasts will be driven more by these successes, or by perceived failures, remains to be seen.

Notably, both Ebola and especially COVID-19 are emerging infectious diseases, for which little prior knowledge exists for informing forecasting models. While the same will be true for future influenza pandemics, we know much more about the epidemiologic parameters realistic for seasonal flu (although, as discussed throughout this chapter, several gaps in our knowledge do remain). Furthermore, we have shown that real-time forecasts are plausible at the country level for several countries, and generally outperform methods based on historical expectance (see Chapter 4); similar results have been found in both US (Shaman et al., 2013) and Australian (Moss et al., 2018) cities. The above examples show that practitioners and policymakers are willing to incorporate forecasts into their decision making, and furthermore are even willing to trust forecasts during high-stakes pandemics. Given this, the outlook for operationalizing real-time forecasts of seasonal influenza outbreaks appears optimistic. Again, the collaborations discussed above will be helpful in making this step, not only because of the increase in meaningful communication concerning forecast results and assumptions, but also because modelers involved in such collaborations may gain better access to data in real-time (Lipsitch et al., 2011).

6.3: Implications beyond forecasting and for other diseases

Beyond forecasting

While the bulk of this dissertation has focused on forecasting, our work has broader implications for the study of influenza. For example, we observed in Chapter 3 that (sub)type

dynamics were not always consistent between European countries. Future studies could use WHO data to assess (sub)type dynamics in countries reporting good-quality data (see Chapter 2). In particular, describing patterns in (sub)type dominance between consecutive outbreaks could improve our understanding of how influenza immunity contributes to transmission patterns, and identifying countries with a tendency to report similar or dissimilar dynamics could improve our understanding of the role of travel.

Influenza seasonality is well-understood in temperate regions, but patterns in the tropics and subtropics remain poorly characterized in many countries (Cummings et al., 2016; Shek & Lee, 2003). In Chapter 2, we find that eighteen countries in tropical and subtropical regions report good-quality data to the WHO, and many of these countries have historical data spanning multiple years. These data could be used to better characterize influenza seasonality, or lack thereof, in these countries, and to compare outbreak patterns between countries. Results could contribute to a better understanding of climatic and other factors that drive influenza outbreaks in tropical regions.

We have explained that the need for influenza forecasts stems partially from the fact that, even when outbreaks in a given location display a clear seasonal pattern, substantial heterogeneities in outbreak timing and intensity are observed. Understanding the drivers behind these heterogeneities could further improve influenza preparedness. Future work should explore how outbreak timing and intensity vary by observed absolute humidity (Shaman et al., 2010), as well as by longer-term climate phenomena such as the El Niño-Southern Oscillation and the North Atlantic Oscillation (Oluwole, 2015; Viboud, Pakdaman, et al., 2004; Zaraket et al., 2008).

Beyond influenza

As discussed throughout this dissertation, syndromic data such as ILI are not specific for influenza. Therefore, countries with viral positivity data for other respiratory infections of public health importance, such as respiratory syncytial virus (RSV), human parainfluenza virus, human metapneumovirus, respiratory adenovirus, and rhinovirus (Pavia, 2011), could use these in conjunction with FluID data to calculate syndromic+ data for a variety of infections. Because there are likely substantial differences in health seeking behavior between viruses based on the severity of clinical disease they cause (Galanti et al., 2020), these data will not necessarily reflect the relative intensity of the different infections. However, such data could still be useful in exploring the seasonality of these pathogens, for example, as long as summer sampling is sufficient. While past work exists exploring the seasonality of most of these pathogens (Bloom-Feshbach et al., 2013, p.; Li et al., 2019; Maykowski et al., 2018; Monto, 2002), particularly RSV, further work is needed. Analyses of the seasonal drivers of most of these pathogens in particular appear to be lacking (Martinez, 2018). Importantly, as discussed earlier in this chapter, within-country drivers likely play a large role in shaping influenza transmission dynamics at the country-level; we may expect key drivers to be even more localized in the case of infections that primarily occur in children, as children are less likely to engage in travel between locations (Apolloni et al., 2013; Charaudeau et al., 2014). These data could also be used to develop forecasting systems for a variety of infections, something that has shown promise for RSV in the US (Reis et al., 2019), and which can improve forecasts of overall ILI (Pei & Shaman, 2020).

The existence of good-quality influenza data, as well as skillful influenza forecasting systems, can be relevant for chronic disease epidemiology, as well. Influenza can increase the risk of cardiovascular disease-related mortality, and was found to be predictive of cardiovascular

mortality, particularly myocardial infarctions, among adults aged 65 and older in New York City (Nguyen et al., 2016). A recent study found that other non-respiratory conditions, in particular sepsis and acute kidney injury, were also common among adults hospitalized with influenza (Chow et al., 2020). The fact that we can skillfully forecast influenza in many countries means that predictions of increased rates of these non-respiratory conditions may also be possible, albeit at the country scale. Influenza forecasts and nowcasts may also help inform practitioners as to when influenza should be considered in patients hospitalized for certain non-respiratory conditions, which may in turn lead to improved rates of antiviral prescription among these patients (Chow et al., 2020).

Conclusion

Influenza causes substantial morbidity and mortality worldwide, both through regular outbreaks and sporadic, potentially severe pandemics (Dawood et al., 2012; Iuliano et al., 2018; Taubenberger & Morens, 2006). Accurate and well-calibrated forecasts have the potential to reduce this morbidity and mortality if used to support public health decision making. In this dissertation, we explored the potential of forecasts in several countries worldwide, with a particular focus on Europe, and highlighted several issues hindering both forecast quality and forecast operationalization. Specifically, countries should continue to invest in improving their influenza surveillance programs, and the WHO should push countries to report more informative denominator data, particularly for epidemiologic surveillance data. Further research is also needed to better understand the drivers of influenza transmission, including climatic drivers, in both temperate and tropical regions. While these improvements will be impactful, we have also highlighted several steps that can be taken to improve forecast performance using the tools and

information currently available. Finally, modelers must work together with public health and medical professionals to develop forecasts that are relevant to public health practice, and to integrate these forecasts into decision making. Since the ultimate goal of forecasting is to improve public health, this last step is particularly crucial.

References

Adam, D. (2020). Special report: The simulations driving the world's response to COVID-19. *Nature*, *580*(7803), 316–318. https://doi.org/10.1038/d41586-020-01003-6

Alonso, W. J., Viboud, C., Simonsen, L., Hirano, E. W., Daufenbach, L. Z., & Miller, M. A. (2007). Seasonality of Influenza in Brazil: A Traveling Wave from the Amazon to the Subtropics. *American Journal of Epidemiology*, *165*(12), 1434–1442. https://doi.org/10.1093/aje/kwm012

Alwan, N. A., Bhopal, R., Burgess, R. A., Colburn, T., Cuevas, L. E., Smith, G. D., Egger, M., Eldridge, S., Gallo, V., Gilthorpe, M. S., Greenhalgh, T., Griffiths, C., Hunter, P. R., Jaffar, S., Jepson, R., Low, N., Martineau, A., McCoy, D., Orcutt, M., … Wilson, A. (2020). Evidence informing the UK's COVID-19 public health response must be transparent. *The Lancet*, *395*(10229), 1036–1037. https://doi.org/10.1016/S0140-6736(20)30667-X

Apolloni, A., Poletto, C., & Colizza, V. (2013). Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic. *BMC Infectious Diseases*, *13*(1), 176. https://doi.org/10.1186/1471-2334-13-176

Bedford, T., Cobey, S., Beerli, P., & Pascual, M. (2010). Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2). *PLoS Pathogens*, *6*(5), e1000918. https://doi.org/10.1371/journal.ppat.1000918

Bedford, T., Riley, S., Barr, I. G., Broor, S., Chadha, M., Cox, N. J., Daniels, R. S., Gunasekaran, C. P., Hurt, A. C., Kelso, A., Klimov, A., Lewis, N. S., Li, X., McCauley, J. W., Odagiri, T., Potdar, V., Rambaut, A., Shu, Y., Skepner, E., … Russell, C. A. (2015). Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, *523*(7559), 217–220. https://doi.org/10.1038/nature14460

Biggerstaff, M., Dahlgren, F. S., Fitzner, J., George, D., Hammond, A., Hall, I., Haw, D., Imai, N., Johansson, M. A., Kramer, S., McCaw, J. M., Moss, R., Pebody, R., Read, J. M., Reed, C., Reich, N. G., Riley, S., Vandemaele, K., Viboud, C., & Wu, J. T. (2019). Coordinating the real-time use of global influenza activity data for better public health planning. *Influenza and Other Respiratory Viruses*, *14*(2), 105–110. https://doi.org/10.1111/irv.12705

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. (2018). Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*. https://doi.org/10.1016/j.epidem.2018.02.003

Bloom-Feshbach, K., Alonso, W. J., Charu, V., Tamerius, J., Simonsen, L., Miller, M. A., & Viboud, C. (2013). Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PLOS ONE*, *8*(2), e54445. https://doi.org/10.1371/journal.pone.0054445

Brooks, L. C., Farrow, D. C., Hyun, S., Tibshirani, R. J., & Rosenfeld, R. (2015). Flexible Modeling of Epidemics with an Empirical Bayes Framework. *PLoS Computational Biology*, *11*(8). https://doi.org/10.1371/journal.pcbi.1004382

Brownstein, J. S., Wolfe, C. J., & Mandl, K. D. (2006). Empirical Evidence for the Effect of Airline Travel on Inter-Regional Influenza Spread in the United States. *PLoS Medicine*, *3*(10), e401. https://doi.org/10.1371/journal.pmed.0030401

Caini, S., Andrade, W., Badur, S., Balmaseda, A., Barakat, A., Bella, A., Bimohuen, A., Brammer, L., Bresee, J., Bruno, A., Castillo, L., Ciblak, M. A., Clara, A. W., Cohen, C., Cutter, J., Daouda, C., de Lozano, C., De Mora, D., Dorji, K., … Global Influenza B Study. (2016). Temporal Patterns of Influenza A and B in Tropical and Temperate Countries: What Are the Lessons for Influenza Vaccination? *PloS One*, *11*(3), e0152310. https://doi.org/10.1371/journal.pone.0152310

CDC. (n.d.-a). *Epidemic Prediction Initiative*. Centers for Disease Control and Prevention. https://predict.cdc.gov/

CDC. (n.d.-b). *FluView Interactive*. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/weekly/fluviewinteractive.htm

Charaudeau, S., Pakdaman, K., & Boëlle, P.-Y. (2014). Commuter Mobility and the Spread of Infectious Diseases: Application to Influenza in France. *PLoS ONE*, *9*(1), e83002. https://doi.org/10.1371/journal.pone.0083002

Chow, E. J., Rolfes, M. A., O'Halloran, A., Alden, N. B., Anderson, E. J., Bennett, N. M., Billing, L., Dufort, E., Kirley, P. D., George, A., Irizarry, L., Kim, S., Lynfield, R., Ryan, P., Schaffner, W., Talbot, H. K., Thomas, A., Yousey-Hindes, K., Reed, C., & Garg, S. (2020). Respiratory and Nonrespiratory Diagnoses Associated With Influenza in Hospitalized Adults. *JAMA Network Open*, *3*(3), e201323. https://doi.org/10.1001/jamanetworkopen.2020.1323

Chowell, G., Viboud, C., Simonsen, L., Merler, S., & Vespignani, A. (2017). Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: Lessons and the way forward. *BMC Medicine*, *15*(1), 42, s12916-017-0811-y. https://doi.org/10.1186/s12916-017-0811-y

Columbia University Mailman School of Public Health. (n.d.). *Influenza Observations and Forecast.* Columbia Prediction of Infectious Diseases. https://cpid.iri.columbia.edu/

Crepey, P., & Barthelemy, M. (2007). Detecting Robust Patterns in the Spread of Epidemics: A Case Study of Influenza in the United States and France. *American Journal of Epidemiology*, *166*(11), 1244–1251. https://doi.org/10.1093/aje/kwm266

Cummings, M. J., Bakamutumaho, B., Kayiwa, J., Byaruhanga, T., Owor, N., Namagambo, B., Wolf, A., Wamala, J. F., Morse, S. S., Lutwama, J. J., & O'Donnell, M. R. (2016). Epidemiologic and Spatiotemporal Characterization of Influenza and Severe Acute

Respiratory Infection in Uganda, 2010-2015. *Annals of the American Thoracic Society*, *13*(12), 2159–2168. https://doi.org/10.1513/AnnalsATS.201607-561OC

Dalziel, B. D., Kissler, S., Gog, J. R., Viboud, C., Bjørnstad, O. N., Metcalf, C. J. E., & Grenfell, B. T. (2018). Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science*, *362*(6410), 75–79. https://doi.org/10.1126/science.aat6030

Dawood, F. S., Iuliano, A. D., Reed, C., Meltzer, M. I., Shay, D. K., Cheng, P.-Y., Bandaranayake, D., Breiman, R. F., Brooks, W. A., Buchy, P., Feikin, D. R., Fowler, K. B., Gordon, A., Hien, N. T., Horby, P., Huang, Q. S., Katz, M. A., Krishnan, A., Lal, R., … Widdowson, M.-A. (2012). Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study. *The Lancet Infectious Diseases*, *12*(9), 687–695. https://doi.org/10.1016/S1473-3099(12)70121-4

de Lusignan, S., Borrow, R., Tripathy, M., Linley, E., Zambon, M., Hoschler, K., Ferreira, F., Andrews, N., Yonova, I., Hriskova, M., Rafi, I., & Pebody, R. (2019). Serological surveillance of influenza in an English sentinel network: Pilot study protocol. *BMJ Open*, *9*(3), e024285. https://doi.org/10.1136/bmjopen-2018-024285

Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., & Sugihara, G. (2016). Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences*, *113*(46), 13081–13086. https://doi.org/10.1073/pnas.1607747113

Driedger, S. M., Cooper, E. J., & Moghadas, S. M. (2014). Developing model-based public health policy through knowledge translation: The need for a 'Communities of Practice.' *Public Health*, *128*(6), 561–567. https://doi.org/10.1016/j.puhe.2013.10.009

ECDC. (2014). *Data quality monitoring and surveillance system evaluation: A handbook of methods and applications.* European Centre for Disease Prevention and Control. http://bookshop.europa.eu/uri?target=EUB:NOTICE:TQ0414829:EN:HTML

ECDC. (2018). *Seasonal influenza vaccination and antiviral use in EU/EEA Member States – Overview of vaccine recommendations for 2017–2018 and vaccination coverage rates for 2015–2016 and 2016–2017 influenza seasons*. ECDC. https://www.ecdc.europa.eu/sites/default/files/documents/seasonal-influenza-antiviral-use-2018.pdf

Enserink, M., & Kupferschmidt, K. (2020). With COVID-19, modeling takes on life and death importance. *Science*, *367*(6485), 1414.2-1415. https://doi.org/10.1126/science.367.6485.1414-b

European Commission, & Statistical Office of the European Union. (2018). *Smarter, greener, more inclusive?: Indicators to support the Europe 2020 strategy: 2018 edition.* http://dx.publications.europa.eu/10.2785/170012

Ewing, A., Lee, E. C., Viboud, C., & Bansal, S. (2016). Contact, travel, and transmission: The impact of winter holidays on influenza dynamics in the United States. *Journal of Infectious Diseases*, jiw642. https://doi.org/10.1093/infdis/jiw642

Farrow, D. C., Brooks, L. C., Hyun, S., Tibshirani, R. J., Burke, D. S., & Rosenfeld, R. (2017). A human judgment approach to epidemiological forecasting. *PLoS Computational Biology*, *13*(3). https://doi.org/10.1371/journal.pcbi.1005248

Flahault, A., Blanchon, T., Dorléans, Y., Toubiana, L., Vibert, J. F., & Valleron, A. J. (2006). Virtual surveillance of communicable diseases: A 20-year experience in France. *Statistical Methods in Medical Research*, *15*(5), 413–421. https://doi.org/10.1177/0962280206071639

Funk, S., Camacho, A., Kucharski, A. J., Lowe, R., Eggo, R. M., & Edmunds, W. J. (2019). Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. *PLOS Computational Biology*, *15*(2), e1006785. https://doi.org/10.1371/journal.pcbi.1006785

Galanti, M., Comito, D., Ligon, C., Lane, B., Matienzo, N., Ibrahim, S., Shittu, A., Tagne, E., Birger, R., Ud-Dean, M., Filip, I., Morita, H., Rabadan, R., Anthony, S., Freyer, G. A., Dayan, P., Shopsin, B., & Shaman, J. (2020). Active surveillance documents rates of clinical care seeking due to respiratory illness. *Influenza and Other Respiratory Viruses*, irv.12753. https://doi.org/10.1111/irv.12753

Geoghegan, J. L., Saavedra, A. F., Duchêne, S., Sullivan, S., Barr, I., & Holmes, E. C. (2018). Continental synchronicity of human influenza virus epidemics despite climactic variation. *PLOS Pathogens*, *14*(1), e1006780. https://doi.org/10.1371/journal.ppat.1006780

George, D. B., Taylor, W., Shaman, J., Rivers, C., Paul, B., O'Toole, T., Johansson, M. A., Hirschman, L., Biggerstaff, M., Asher, J., & Reich, N. G. (2019). Technology to advance infectious disease forecasting for outbreak management. *Nature Communications*, *10*(1), 3932. https://doi.org/10.1038/s41467-019-11901-7

Harper, G. J. (1961). Airborne micro-organisms: Survival tests with four viruses. *Journal of Hygiene*, *59*(04), 479–486. https://doi.org/10.1017/S0022172400039176

Haynes, S. (2020, March 6). As Coronavirus Spreads, So Does Xenophobia and Anti-Asian Racism. *Time*. https://time.com/5797836/coronavirus-racism-stereotypes-attacks/

Heaney, A., Little, E., Ng, S., & Shaman, J. (2016). Meteorological variability and infectious disease in Central Africa: A review of meteorological data quality: Meteorology and infectious disease in C. Africa. *Annals of the New York Academy of Sciences*, *1382*(1), 31–43. https://doi.org/10.1111/nyas.13090

Hirve, S., Newman, L. P., Paget, J., Azziz-Baumgartner, E., Fitzner, J., Bhat, N., Vandemaele, K., & Zhang, W. (2016). Influenza Seasonality in the Tropics and Subtropics – When to Vaccinate? *PLOS ONE*, *11*(4), e0153003. https://doi.org/10.1371/journal.pone.0153003

Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., Cohen, C., Gran, J. M., Schanzer, D., Cowling, B. J., Wu, P., Kyncl, J., Ang, L. W., Park, M., Redlberger-Fritz, M., Yu, H., Espenhain, L., Krishnan, A., Emukule, G., … Mustaquim, D. (2018). Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *The Lancet*, *391*(10127), 1285–1300. https://doi.org/10.1016/S0140-6736(17)33293-2

Kandula, S., Hsu, D., & Shaman, J. (2017). Subregional Nowcasts of Seasonal Influenza Using Search Trends. *Journal of Medical Internet Research*, *19*(11), e370. https://doi.org/10.2196/jmir.7486

Kandula, S., Pei, S., & Shaman, J. (2019). Improved forecasts of influenza-associated hospitalization rates with Google Search Trends. *Journal of The Royal Society Interface*, *16*(155), 20190080. https://doi.org/10.1098/rsif.2019.0080

Kandula, S., Yamana, T., Pei, S., Yang, W., Morita, H., & Shaman, J. (2018). Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *Journal of The Royal Society Interface*, *15*(144), 20180174. https://doi.org/10.1098/rsif.2018.0174

Kandula, S., Yang, W., & Shaman, J. (2017). Type- and Subtype-Specific Influenza Forecast. *American Journal of Epidemiology*, *185*(5), 395–402. https://doi.org/10.1093/aje/kww211

Katz, M. A., Schoub, B. D., Heraud, J. M., Breiman, R. F., Njenga, M. K., & Widdowson, M.-A. (2012). Influenza in Africa: Uncovering the Epidemiology of a Long-Overlooked Disease. *Journal of Infectious Diseases*, *206*(suppl 1), S1–S4. https://doi.org/10.1093/infdis/jis548

Keeling, M. J. (2005). Models of foot-and-mouth disease. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1569), 1195–1202. https://doi.org/10.1098/rspb.2004.3046

Khan, M. S., Osei-Kofi, A., Omar, A., Kirkbride, H., Kessel, A., Abbara, A., Heymann, D., Zumla, A., & Dar, O. (2016). Pathogens, prejudice, and politics: The role of the global health community in the European refugee crisis. *The Lancet Infectious Diseases*, *16*(8), e173–e177. https://doi.org/10.1016/S1473-3099(16)30134-7

Kusunose, Y., & Mahmood, R. (2016). Imperfect forecasts and decision making in agriculture. *Agricultural Systems*, *146*, 103–110. https://doi.org/10.1016/j.agsy.2016.04.006

Lampos, V., Miller, A. C., Crossan, S., & Stefansen, C. (2015). Advances in nowcasting influenza-like illness rates using search query logs. *Scientific Reports*, *5*(1), 12760. https://doi.org/10.1038/srep12760

Lau, M. S. Y., Cowling, B. J., Cook, A. R., & Riley, S. (2015). Inferring influenza dynamics and control in households. *Proceedings of the National Academy of Sciences*, *112*(29), 9094–9099. https://doi.org/10.1073/pnas.1423339112

Lee, B. Y., Haidari, L. A., & Lee, M. S. (2013). Modelling during an emergency: The 2009 H1N1 influenza pandemic. *Clinical Microbiology and Infection*, *19*(11), 1014–1022. https://doi.org/10.1111/1469-0691.12284

Lemey, P., Rambaut, A., Drummond, A. J., & Suchard, M. A. (2009). Bayesian Phylogeography Finds Its Roots. *PLoS Computational Biology*, *5*(9), e1000520. https://doi.org/10.1371/journal.pcbi.1000520

Li, Y., Reeves, R. M., Wang, X., Bassat, Q., Brooks, W. A., Cohen, C., Moore, D. P., Nunes, M., Rath, B., Campbell, H., Nair, H., Acacio, S., Alonso, W. J., Antonio, M., Ayora Talavera, G., Badarch, D., Baillie, V. L., Barrera-Badillo, G., Bigogo, G., … Zar, H. J. (2019). Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: A systematic analysis. *The Lancet Global Health*, *7*(8), e1031–e1045. https://doi.org/10.1016/S2214-109X(19)30264-5

Lipsitch, M., Finelli, L., Heffernan, R. T., Leung, G. M., & Redd, S. C. (2011). Improving the Evidence Base for Decision Making During a Pandemic: The Example of 2009 Influenza A/H1N. *Biosecurity and Bioterrorism*, *9*(2), 28.

Lofgren, E. T., Halloran, M. E., Rivers, C. M., Drake, J. M., Porco, T. C., Lewis, B., Yang, W., Vespignani, A., Shaman, J., Eisenberg, J. N. S., Eisenberg, M. C., Marathe, M., Scarpino, S. V., Alexander, K. A., Meza, R., Ferrari, M. J., Hyman, J. M., Meyers, L. A., & Eubank, S. (2014). Opinion: Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences*, *111*(51), 18095–18096. https://doi.org/10.1073/pnas.1421551111

Lowen, A. C., Mubareka, S., Steel, J., & Palese, P. (2007). Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. *PLoS Pathogens*, *3*(10). https://doi.org/10.1371/journal.ppat.0030151

Lowen, A. C., Steel, J., Mubareka, S., & Palese, P. (2008). High Temperature (30°C) Blocks Aerosol but Not Contact Transmission of Influenza Virus. *Journal of Virology*, *82*(11), 5650–5652. https://doi.org/10.1128/JVI.00325-08

Lu, F. S., Hattab, M. W., Clemente, C. L., Biggerstaff, M., & Santillana, M. (2019). Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches. *Nature Communications*, *10*(1), 147. https://doi.org/10.1038/s41467-018-08082-0

Lucyk, K., Simmonds, K. A., Lorenzetti, D. L., Drews, S. J., Svenson, L. W., & Russell, M. L. (2019). The association between influenza vaccination and socioeconomic status in high income countries varies by the measure used: A systematic review. *BMC Medical Research Methodology*, *19*(1), 153. https://doi.org/10.1186/s12874-019-0801-1

Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., Fernandez, D., Greene, S. K., Kipshidze, N., Liu, L., Mgbere, O., McHugh, L. A., Myers, J. F., Siniscalchi, A., Sullivan, A. D., West, N., Johansson, M. A., & Biggerstaff, M. (2019). Applying infectious disease forecasting to public health: A path forward using

influenza forecasting examples. *BMC Public Health*, *19*(1), 1659. https://doi.org/10.1186/s12889-019-7966-8

Martinez, M. E. (2018). The calendar of epidemics: Seasonal cycles of infectious diseases. *PLOS Pathogens*, *14*(11), e1007327. https://doi.org/10.1371/journal.ppat.1007327

Maykowski, P., Smithgall, M., Zachariah, P., Oberhardt, M., Vargas, C., Reed, C., Demmer, R. T., Stockwell, M. S., & Saiman, L. (2018). Seasonality and clinical impact of human parainfluenza viruses. *Influenza and Other Respiratory Viruses*, *12*(6), 706–716. https://doi.org/10.1111/irv.12597

Mayoral, J. M., Alonso, J., Garín, O., Herrador, Z., Astray, J., Baricot, M., Castilla, J., Cantón, R., Castro, A., Delgado-Rodríguez, M., Ferri, A., Godoy, P., González-Candelas, F., Martín, V., Pumarola, T., Quintana, J. M., Soldevila, N., Tamames, S., Domínguez, A., & CIBERESP Cases and Controls in Pandemic Influenza Working Group, Spain. (2013). Social factors related to the clinical severity of influenza cases in Spain during the A (H1N1) 2009 virus pandemic. *BMC Public Health*, *13*, 118. https://doi.org/10.1186/1471-2458-13-118

Meltzer, M. I., Atkins, C. Y., Santibanez, S., Knust, B., Petersen, B. W., Ervin, E. D., Nichol, S. T., Damon, I. K., Washington, M. L., & Centers for Disease Control and Prevention (CDC). (2014). Estimating the future number of cases in the Ebola epidemic—Liberia and Sierra Leone, 2014-2015. *MMWR Supplements*, *63*(3), 1–14.

Metcalf, C. J. E., Edmunds, W. J., & Lessler, J. (2015). Six challenges in modelling for public health policy. *Epidemics*, *10*, 93–96. https://doi.org/10.1016/j.epidem.2014.08.008

Metzger, K. B., Hajat, A., Crawford, M., & Mostashari, F. (2004). How many illnesses does one emergency department visit represent? Using a population-based telephone survey to estimate the syndromic multiplier. *MMWR Supplements*, *53*, 106–111.

Milwid, R., Steriu, A., Arino, J., Heffernan, J., Hyder, A., Schanzer, D., Gardner, E., Haworth-Brockman, M., Isfeld-Kiely, H., Langley, J. M., & Moghadas, S. M. (2016). Toward Standardizing a Lexicon of Infectious Disease Modeling Terms. *Frontiers in Public Health*, *4*. https://doi.org/10.3389/fpubh.2016.00213

Modeling for Public Health (mod4PH). (n.d.). LinkedIn. https://www.linkedin.com/groups/6787233/

Moghadas, S. M., Haworth-Brockman, M., Isfeld-Kiely, H., & Kettner, J. (2015). Improving Public Health Policy through Infection Transmission Modelling: Guidelines for Creating a Community of Practice. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *26*(4), 191–195. https://doi.org/10.1155/2015/274569

Mohammadi, D. (2015). Ebola vaccine trials back on track. *The Lancet*, *385*(9964), 214–215. https://doi.org/10.1016/S0140-6736(15)60035-6

Monto, A. S. (2002). The seasonality of rhinovirus infections and its implications for clinical recognition. *Clinical Therapeutics*, *24*(12), 1987–1997. https://doi.org/10.1016/S0149-2918(02)80093-5

Morgan, O. (2019). How decision makers can use quantitative approaches to guide outbreak responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1776), 20180365. https://doi.org/10.1098/rstb.2018.0365

Morita, H., Kramer, S., Heaney, A., Gil, H., & Shaman, J. (2018). Influenza forecast optimization when using different surveillance data types and geographic scale. *Influenza and Other Respiratory Viruses*. https://doi.org/10.1111/irv.12594

Morris, S. E., Freiesleben de Blasio, B., Viboud, C., Wesolowski, A., Bjørnstad, O. N., & Grenfell, B. T. (2018). Analysis of multi-level spatial data reveals strong synchrony in seasonal influenza epidemics across Norway, Sweden, and Denmark. *PLOS ONE*, *13*(5), e0197519. https://doi.org/10.1371/journal.pone.0197519

Moss, R., Fielding, J. E., Franklin, L. J., Stephens, N., McVernon, J., Dawson, P., & McCaw, J. M. (2018). Epidemic forecasts as a tool for public health: Interpretation and (re)calibration. *Australian and New Zealand Journal of Public Health*, *42*(1), 69–76. https://doi.org/10.1111/1753-6405.12750

Moss, R., Zarebski, A., Carlson, S., & McCaw, J. (2019). Accounting for Healthcare-Seeking Behaviours and Testing Practices in Real-Time Influenza Forecasts. *Tropical Medicine and Infectious Disease*, *4*(1), 12. https://doi.org/10.3390/tropicalmed4010012

Moss, R., Zarebski, A., Dawson, P., & McCAW, J. M. (2017). Retrospective forecasting of the 2010–2014 Melbourne influenza seasons using multiple surveillance systems. *Epidemiology & Infection*, *145*(1), 156–169. https://doi.org/10.1017/S0950268816002053

Moss, R., Zarebski, A. E., Dawson, P., Franklin, L. J., Birrell, F. A., & McCaw, J. M. (2019). Anatomy of a seasonal influenza epidemic forecast. *Communicable Diseases Intelligence*, *43*. https://doi.org/10.33321/cdi.2019.43.7

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine*, *5*(3), e74. https://doi.org/10.1371/journal.pmed.0050074

Muscatello, D. J., Chughtai, A. A., Heywood, A., Gardner, L. M., Heslop, D. J., & MacIntyre, C. R. (2017). Translation of Real-Time Infectious Disease Modeling into Routine Public Health Practice. *Emerging Infectious Diseases*, *23*(5). https://doi.org/10.3201/eid2305.161720

Ng, S., & Gordon, A. (2015). Influenza Burden and Transmission in the Tropics. *Current Epidemiology Reports*, *2*(2), 89–100. https://doi.org/10.1007/s40471-015-0038-4

Nguyen, J. L., Yang, W., Ito, K., Matte, T. D., Shaman, J., & Kinney, P. L. (2016). Seasonal Influenza Infections and Cardiovascular Disease Mortality. *JAMA Cardiology*, *1*(3), 274. https://doi.org/10.1001/jamacardio.2016.0433

Oluwole, O. S. A. (2015). Seasonal Influenza Epidemics and El Niños. *Frontiers in Public Health*, *3*. https://doi.org/10.3389/fpubh.2015.00250

Ortiz, J. R., Sotomayor, V., Uez, O. C., Oliva, O., Bettels, D., McCarron, M., Bresee, J. S., & Mounts, A. W. (2009). Strategy to Enhance Influenza Surveillance Worldwide. *Emerging Infectious Diseases*, *15*(8), 1271–1278. https://doi.org/10.3201/eid1508.081422

Osthus, D., Gattiker, J., Priedhorsky, R., & Del Valle, S. Y. (2019). Dynamic Bayesian Influenza Forecasting in the United States with Hierarchical Discrepancy (with Discussion). *Bayesian Analysis*, *14*(1), 261–312. https://doi.org/10.1214/18-BA1117

Paget, J., Marquet, R., Meijer, A., & van der Velden, K. (2007). Influenza activity in Europe during eight seasons (1999–2007): An evaluation of the indicators used to measure activity and an assessment of the timing, length and course of peak activity (spread) across Europe. *BMC Infectious Diseases*, *7*(1), 141. https://doi.org/10.1186/1471-2334-7-141

Patterson Ross, Z., Komadina, N., Deng, Y.-M., Spirason, N., Kelly, H. A., Sullivan, S. G., Barr, I. G., & Holmes, E. C. (2015). Inter-Seasonal Influenza is Characterized by Extended Virus Transmission and Persistence. *PLOS Pathogens*, *11*(6), e1004991. https://doi.org/10.1371/journal.ppat.1004991

Paul, M. J., Dredze, M., & Broniatowski, D. (2014). Twitter Improves Influenza Forecasting. *PLoS Currents*. https://doi.org/10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117

Pavia, A. T. (2011). Viral Infections of the Lower Respiratory Tract: Old Viruses, New Viruses, and the Role of Diagnosis. *Clinical Infectious Diseases*, *52*(Supplement 4), S284–S289. https://doi.org/10.1093/cid/cir043

Pei, S., Cane, M. A., & Shaman, J. (2019). Predictability in process-based ensemble forecast of influenza. *PLOS Computational Biology*, *15*(2), e1006783. https://doi.org/10.1371/journal.pcbi.1006783

Pei, S., Kandula, S., Yang, W., & Shaman, J. (2018). Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences*, 201708856. https://doi.org/10.1073/pnas.1708856115

Pei, S., & Shaman, J. (2017). Counteracting structural errors in ensemble forecast of influenza outbreaks. *Nature Communications*, *8*(1), 925. https://doi.org/10.1038/s41467-017-01033-1

Pei, S., & Shaman, J. (2020). Aggregating forecasts of multiple respiratory pathogens supports more accurate forecasting of influenza-like illness. [Under Review.]

Polansky, L. S., Outin-Blenman, S., & Moen, A. C. (2016). Improved Global Capacity for Influenza Surveillance. *Emerging Infectious Diseases*, *22*(6), 993–1001. https://doi.org/10.3201/eid2206.151521

Radin, J. M., Katz, M. A., Tempia, S., Talla Nzussouo, N., Davis, R., Duque, J., Adedeji, A., Adjabeng, M. J., Ampofo, W. K., Ayele, W., Bakamutumaho, B., Barakat, A., Cohen, A. L., Cohen, C., Dalhatu, I. T., Daouda, C., Dueger, E., Francisco, M., Heraud, J.-M., … Widdowson, M.-A. (2012). Influenza Surveillance in 15 Countries in Africa, 2006–2010. *The Journal of Infectious Diseases*, *206*(suppl_1), S14–S21. https://doi.org/10.1093/infdis/jis606

Ray, E. L., & Reich, N. G. (2018). Prediction of infectious disease epidemics via weighted density ensembles. *PLOS Computational Biology*, *14*(2), e1005910. https://doi.org/10.1371/journal.pcbi.1005910

Reis, J., Yamana, T., Kandula, S., & Shaman, J. (2019). Superensemble forecast of respiratory syncytial virus outbreaks at national, regional, and state levels in the United States. *Epidemics*, *26*, 1–8. https://doi.org/10.1016/j.epidem.2018.07.001

Rivers, C., Chretien, J.-P., Riley, S., Pavlin, J. A., Woodward, A., Brett-Major, D., Maljkovic Berry, I., Morton, L., Jarman, R. G., Biggerstaff, M., Johansson, M. A., Reich, N. G., Meyer, D., Snyder, M. R., & Pollett, S. (2019). Using "outbreak science" to strengthen the use of models during epidemics. *Nature Communications*, *10*(1), 3102. https://doi.org/10.1038/s41467-019-11067-2

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J., Hurt, A. C., de Jong, J. C., Kelso, A., Klimov, A. I., Kageyama, T., Komadina, N., Lapedes, A. S., Lin, Y. P., Mosterin, A., Obuchi, M., … Smith, D. J. (2008). The Global Circulation of Seasonal Influenza A (H3N2) Viruses. *Science*, *320*(5874), 340–346. https://doi.org/10.1126/science.1154137

Samaras, L., García-Barriocanal, E., & Sicilia, M.-A. (2017). Syndromic Surveillance Models Using Web Data: The Case of Influenza in Greece and Italy Using Google Trends. *JMIR Public Health and Surveillance*, *3*(4), e90. https://doi.org/10.2196/publichealth.8015

Semenza, J. C. (2010). Strategies to intervene on social determinants of infectious diseases. *Eurosurveillance*, *15*(27). https://doi.org/10.2807/ese.15.27.19611-en

Shaman, J., & Kandula, S. (2015). Improved Discrimination of Influenza Forecast Accuracy Using Consecutive Predictions. *PLoS Currents*. https://doi.org/10.1371/currents.outbreaks.8a6a3df285af7ca973fab4b22e10911e

Shaman, J., Kandula, S., Yang, W., & Karspeck, A. (2017). The use of ambient humidity conditions to improve influenza forecast. *PLOS Computational Biology*, *13*(11), e1005844. https://doi.org/10.1371/journal.pcbi.1005844

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3837

Shaman, J., & Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3243–3248. https://doi.org/10.1073/pnas.0806852106

Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T., & Lipsitch, M. (2010). Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States. *PLoS Biology*, *8*(2), e1000316. https://doi.org/10.1371/journal.pbio.1000316

Shek, L. P.-C., & Lee, B.-W. (2003). Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric Respiratory Reviews*, *4*(2), 105–111. https://doi.org/10.1016/S1526-0542(03)00024-1

Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. (2008). Obstacles to High-Dimensional Particle Filtering. *Monthly Weather Review*, *136*(12), 4629–4640. https://doi.org/10.1175/2008MWR2529.1

Soebiyanto, R. P., Adimi, F., & Kiang, R. K. (2010). Modeling and Predicting Seasonal Influenza Transmission in Warm Regions Using Climatological Parameters. *PLoS ONE*, *5*(3), e9450. https://doi.org/10.1371/journal.pone.0009450

Stop the coronavirus stigma now. (2020). *Nature*, *580*(7802), 165–165. https://doi.org/10.1038/d41586-020-01009-0

Su, K., Xu, L., Li, G., Ruan, X., Li, X., Deng, P., Li, X., Li, Q., Chen, X., Xiong, Y., Lu, S., Qi, L., Shen, C., Tang, W., Rong, R., Hong, B., Ning, Y., Long, D., Xu, J., … Li, Y. (2019). Forecasting influenza activity using self-adaptive AI model and multi-source data in Chongqing, China. *EBioMedicine*, *47*, 284–292. https://doi.org/10.1016/j.ebiom.2019.08.024

Tamerius, J., Nelson, M. I., Zhou, S. Z., Viboud, C., Miller, M. A., & Alonso, W. J. (2011). Global Influenza Seasonality: Reconciling Patterns across Temperate and Tropical Regions. *Environmental Health Perspectives*, *119*(4), 439–445. https://doi.org/10.1289/ehp.1002383

Tamerius, J., Shaman, J., Alonso, W. J., Bloom-Feshbach, K., Uejio, C. K., Comrie, A., & Viboud, C. (2013). Environmental Predictors of Seasonal Influenza Epidemics across Temperate and Tropical Climates. *PLoS Pathogens*, *9*(3), e1003194. https://doi.org/10.1371/journal.ppat.1003194

Tamerius, J., Viboud, C., Shaman, J., & Chowell, G. (2015). Impact of School Cycles and Environmental Forcing on the Timing of Pandemic Influenza Activity in Mexican States, May-December 2009. *PLOS Computational Biology*, *11*(8), e1004337. https://doi.org/10.1371/journal.pcbi.1004337

Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: The mother of all pandemics. *Rev Biomed*, *17*, 69–79.

The Epidemic Prediction Initiative. (n.d.). HHS.Gov. https://www.hhs.gov/cto/projects/the-epidemic-prediction-initiative/index.html

The ring vaccination trial: A novel cluster randomised controlled trial design to evaluate vaccine efficacy and effectiveness during outbreaks, with special reference to Ebola. (2015). *BMJ*, h3740. https://doi.org/10.1136/bmj.h3740

Valdivia, A., López-Alcalde, J., Vicente, M., Pichiule, M., Ruiz, M., & Ordobas, M. (2010). Monitoring influenza activity in Europe with Google Flu Trends: Comparison with the findings of sentinel physician networks – results for 2009-10. *Eurosurveillance*, *15*(29). https://doi.org/10.2807/ese.15.29.19621-en

Valleron, A. J., Bouvet, E., Garnerin, P., Ménarès, J., Heard, I., Letrait, S., & Lefaucheux, J. (1986). A computer network for the surveillance of communicable diseases: The French experiment. *American Journal of Public Health*, *76*(11), 1289–1292. https://doi.org/10.2105/AJPH.76.11.1289

Van Kerkhove, M. D., & Ferguson, N. M. (2012). Epidemic and intervention modelling – a scientific rationale for policy decisions? Lessons from the 2009 influenza pandemic. *Bulletin of the World Health Organization*, *90*(4), 306–310. https://doi.org/10.2471/BLT.11.097949

Vandendijck, Y., Faes, C., & Hens, N. (2013). Eight Years of the Great Influenza Survey to Monitor Influenza-Like Illness in Flanders. *PLoS ONE*, *8*(5), e64156. https://doi.org/10.1371/journal.pone.0064156

Viboud, C., Alonso, W. J., & Simonsen, L. (2006). Influenza in Tropical Regions. *PLoS Medicine*, *3*(4), e89. https://doi.org/10.1371/journal.pmed.0030089

Viboud, C., Bjørnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., & Grenfell, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, *312*(5772), 447–451.

Viboud, C., Boëlle, P.-Y., Carrat, F., Valleron, A.-J., & Flahault, A. (2003). Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, *158*(10), 996–1006.

Viboud, C., Boëlle, P.-Y., Cauchemez, S., Lavenu, A., Valleron, A.-J., Flahault, A., & Carrat, F. (2004). Risk factors of influenza transmission in households. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, *54*(506), 684–689.

Viboud, C., Pakdaman, K., Boëlle, P., Wilson, M. L., Myers, M. F., Valleron, A.-J., & Flahault, A. (2004). Association of influenza epidemics with global climate variability. *European Journal of Epidemiology*, *19*(11), 1055–1059. https://doi.org/10.1007/s10654-004-2450-9

Wang, Z., Chakraborty, P., Mekaru, S. R., Brownstein, J. S., Ye, J., & Ramakrishnan, N. (2015). Dynamic Poisson Autoregression for Influenza-Like-Illness Case Count Prediction. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*, 1285–1294. https://doi.org/10.1145/2783258.2783291

WHO. (2010). *Surveillance Recommendations for Member States in the Post Pandemic Period*. World Health Organization. https://www.who.int/csr/resources/publications/swineflu/surveillance_post_pandemic_20100812/en/

WHO. (2011). *Manual for the laboratory diagnosis and virological surveillance of influenza*. World Health Organization. https://www.who.int/influenza/gisrs_laboratory/manual_diagnosis_surveillance_influenza/en/

WHO. (2014). *Global Epidemiological Surveillance Standards for Influenza*. World Health Organization. https://www.who.int/influenza/resources/documents/influenza_surveillance_manual/en/

WHO. (2015). *A manual for estimating disease burden associated with seasonal influenza*. http://apps.who.int/iris/bitstream/10665/178801/1/9789241549301_eng.pdf?ua=1

WHO. (2017). *WHO Public Health Research Agenda for Influenza (2017 Update)*. https://www.who.int/influenza/resources/research/publication_research_agenda_2017/en/

WHO, & ECDC. (2019). *Influenza Surveillance: Country, Territory and Area Profiles 2019*. World Health Organization Regional Office for Europe. http://www.euro.who.int/__data/assets/pdf_file/0016/402082/InfluenzaSurveillanceProfiles_2019_en.pdf

Yaari, R., Katriel, G., Stone, L., Mendelson, E., Mandelboim, M., & Huppert, A. (2016). Model-based reconstruction of an epidemic using multiple datasets: Understanding influenza A/H1N1 pandemic dynamics in Israel. *Journal of The Royal Society Interface*, *13*(116), 20160099. https://doi.org/10.1098/rsif.2016.0099

Yamana, T. K., Kandula, S., & Shaman, J. (2017). Individual versus superensemble forecasts of seasonal influenza outbreaks in the United States. *PLOS Computational Biology*, *13*(11), e1005801. https://doi.org/10.1371/journal.pcbi.1005801

Yang, W., Cummings, M. J., Bakamutumaho, B., Kayiwa, J., Owor, N., Namagambo, B., Byaruhanga, T., Lutwama, J. J., O'Donnell, M. R., & Shaman, J. (2018). Transmission dynamics of influenza in two major cities of Uganda. *Epidemics*, *24*, 43–48. https://doi.org/10.1016/j.epidem.2018.03.002

Yang, W., Elankumaran, S., & Marr, L. C. (2012). Relationship between Humidity and Influenza A Viability in Droplets and Implications for Influenza's Seasonality. *PLoS ONE*, *7*(10), e46789. https://doi.org/10.1371/journal.pone.0046789

Yang, W., Karspeck, A., & Shaman, J. (2014). Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLoS Computational Biology*, *10*(4), e1003583. https://doi.org/10.1371/journal.pcbi.1003583

Yang, W., Olson, D. R., & Shaman, J. (2016). Forecasting Influenza Outbreaks in Boroughs and Neighborhoods of New York City. *PLoS Computational Biology*, *12*(11), e1005201.

Yazdanbakhsh, M., & Kremsner, P. G. (2009). Influenza in Africa. *PLoS Medicine*, *6*(12), e1000182. https://doi.org/10.1371/journal.pmed.1000182

Zaraket, H., Saito, R., Tanabe, N., Taniguchi, K., & Suzuki, H. (2008). Association of early annual peak influenza activity with El Niño southern oscillation in Japan. *Influenza and Other Respiratory Viruses*, *2*(4), 127–130. https://doi.org/10.1111/j.1750-2659.2008.00047.x

Zimmer, C., Leuba, S. I., Yaesoubi, R., & Cohen, T. (2018). Use of daily Internet search query data improves real-time projections of influenza epidemics. *Journal of The Royal Society Interface*, *15*(147), 20180220. https://doi.org/10.1098/rsif.2018.0220