Looking for evidence of a high burden of COVID-19 in the United States from influenza-like illness data

Nicholas G. Reich, Evan L. Ray, Graham C. Gibson, Estee Cramer, Caitlin M. Rivers

2020-03-20 20:58:33 CET

Introduction

In December 2019, an outbreak of a novel, the SARS-CoV-2 coronavirus was detected in Wuhan, China. In the intervening weeks, case counts have grown substantially. As of this writing, there are over 234,000 confirmed cases globally and at least 9,840 deaths from what is currently named COVID-19 [1]. It is now understood that the virus transmits efficiently from person to person, with R0 estimates above 2 and perhaps as high as 3.7 [2, 3].

Community transmission is now ongoing in many locations in the United States. Emerging phylogenetic data suggest that sequenced cases to date in the United States and globally share a common ancestor between mid-November and mid-December 2019. However, due to delays in making widespread diagnostic testing available, the burden of COVID-19 in the United States is not well understood. As an effort to understand trends in people seeking care for respiratory symptoms, we analyze publicly available data on influenza-like illness in the US. Specifically, we compare the proportion of weighted influenza like illness (wILI) that tests negative for influenza during the 2019-2020 flu season to trends from previous seasons. If it were the case that SARS-CoV-2 were causing widespread disease in the United States, we might expect to see in recent weeks a higher fraction of ILI specimens that test negative for influenza compared to the same time in past seasons.

Methods

Data

We downloaded publicly available ILINet and WHO-NREVSS data for US Health and Human Services (HHS) regions (Figure 1) and states.

From the ILINet dataset, we downloaded weighted influenza-like illness (wILI), which measures the percentage of doctor's office visits at sentinel providers that had the primary complaint of fever plus an additional influenza-like symptom (cough, and/or sore throat). For the WHO-NREVSS data, we obtained the total number of specimens tested by participating clinical laboratories, as well as the percent of those specimens that tested positive for influenza. We used 23 seasons of data for HHS regions, beginning with the 1997/1998 season, and 10 seasons of data for states, beginning with the 2010/2011 season. All data sources are available at the weekly time-scale, defined as using the MMWR week standard used by the CDC.

The code used to produce this report is available on GitHub at https://github.com/reichlab/ncov.

Influenza-like illness not attributable to influenza

One possible measure of influenza illness not attributable to influenza (ILI-) can be calculated as follows:

$$\text{ILI-} = (1 - \text{proportion of tests positive for influenza}) \times \text{wILI}$$

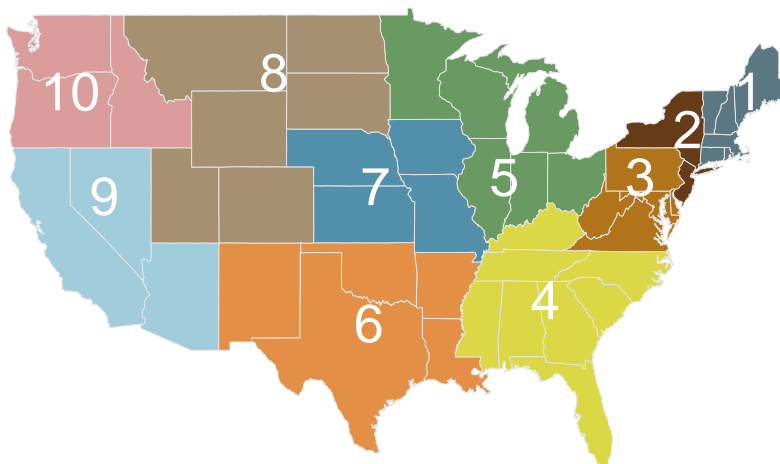US Health and Human Services Regions

Figure 1: US HHS Regions are made up of groups of states.

It is important to note that reported wILI can vary substantially due to differences in the types of health care providers reporting into ILINet. Therefore, some increases in reported wILI from one season to another may be driven in part by changes in provider type make up. An approximate way to adjust for this is by dividing reported wILI by the baseline for a given region and season. Baselines for HHS regions are provided by the CDC. Baselines for states are calculated as the average of the first two weekly ILI observations for a given season, thinking that this adjusts for any systematic adjustments to the provider mix in each season. These baselines enable the following calculation of a **r**elative ILI-.

$$\text{rILI-} = (1 - \text{proportion of tests positive for influenza}) \times \frac{\text{wILI}}{\text{baseline level for ILI}}$$

Measuring anomalies in ILI- during a season

We developed a metric to measure the degree to which a given ILI- observation is significantly higher or lower than expectation, based on past trends at similar times of the year. For each region and season-week, we averaged observations from the past seasons (22 seasons for regions, 9 for states) for the given season week and one season week on either side and calculated the standard deviation based on these same observations. We then computed "z-scores" as the number of standard deviations above or below the average a particular rILI- observation is:

$$Z = \frac{\text{rILI-} - \overline{\text{rILI-}}}{sd\text{rILI-}}$$

Results & Discussion

This report uses data downloaded on March 20, 2020, with data reported through March 14, 2020.

Regional-level analyses

We plotted ILI- and rILI- as a function of the week within each flu season and stratified by region (Figure 2). Additionally, we plotted the 2019/2020 Z-scores for all regions as a function of week of season (Figure 3).

In the last weeks of 2019 and first weeks of 2020, the observations of ILI burden due to non-influenza pathogens (rILI-) have been, relative to what has been observed in the past 22 seasons, above the seasonal

average. More recently, the measures of rILI- also have risen to more than 3 standard deviations above the average in some regions (Figure 3).

These results suggest that there may be enough COVID-19 circulating in the United States to be detectable in the influenza-like illness surveillance system. However, it is hard to determine this conclusively, as we have not performed an exhaustive analysis about what other pathogens were or were not circulating in those past seasons. Also, media attention could also drive more individuals with mild influenza-like illness symptoms to seek care than usual even in the absence of widespread COVID-19 transmission in the US. If these additional individuals seeking care were more likely to have an illness not caused by influenza, then this could also drive up the rILI- metric.

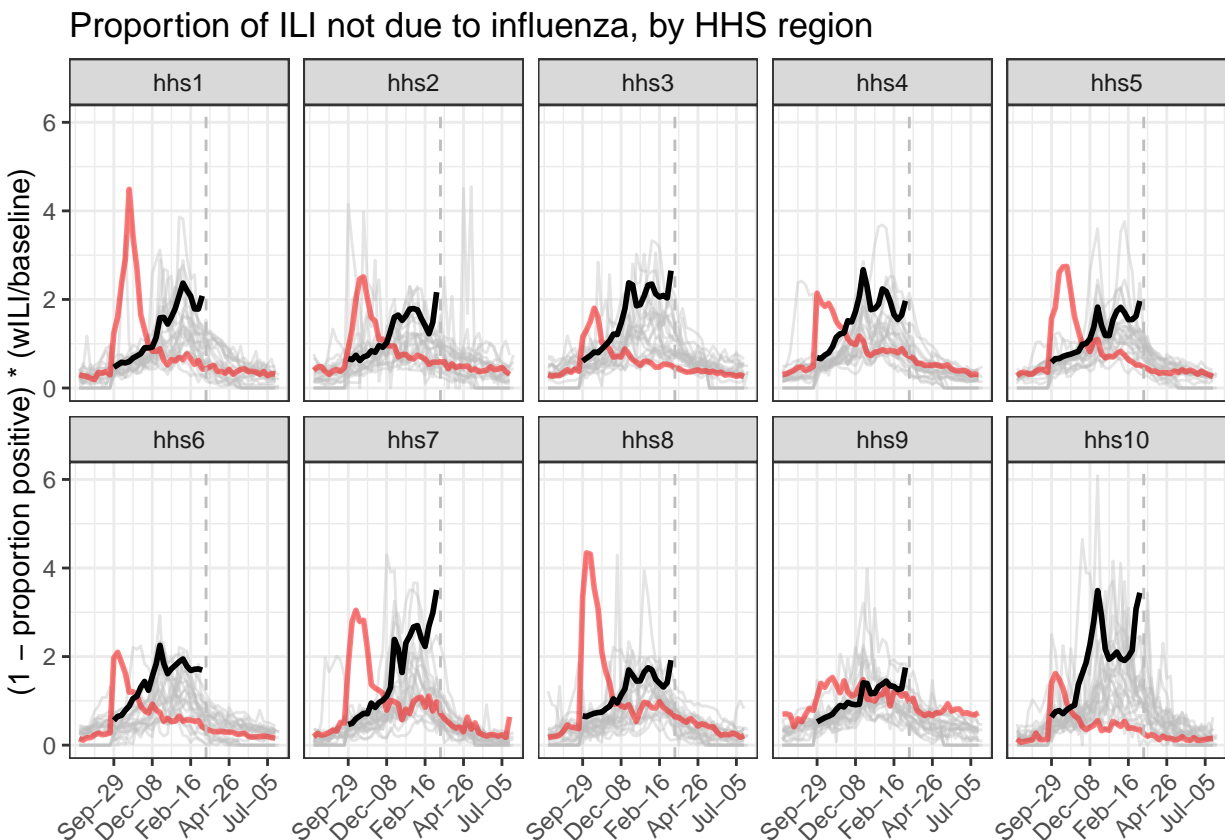## Proportion of ILI not due to influenza, by HHS region

Figure 2: US HHS Regions plots showing rILI- values since the 1997/1998 season (grey lines) and the 2019/2020 season (dark black line). The line highlighted in red is the 2009/2010 H1N1 pandemic season. The dates on the x-axis correspond to the dates for the 2019/2020 season, with previous seasons lined up approximately the same time by week. The vertical dashed line shows the date at which this plot was generated. The small gap between the current season's data and the line indicates the lag in ILI reporting, typically one week.

State-level analyses

The z-score calculations show six states with systematically higher than average observations (Figure 4). We think many of these may be spurious, due to other systematic differences in reporting for this season. Although at this time we do not have specific data to support or refute this hypothesis.

We note an increased signal in Missouri, Pennsylvania, Arizona, California, Massachusetts, and West Virginia (Figure 5). California and Washington are known to have community transmission of SARS-CoV-2, but from this analysis alone we are not able to determine whether that is contributing to the observed signal in rILI-.
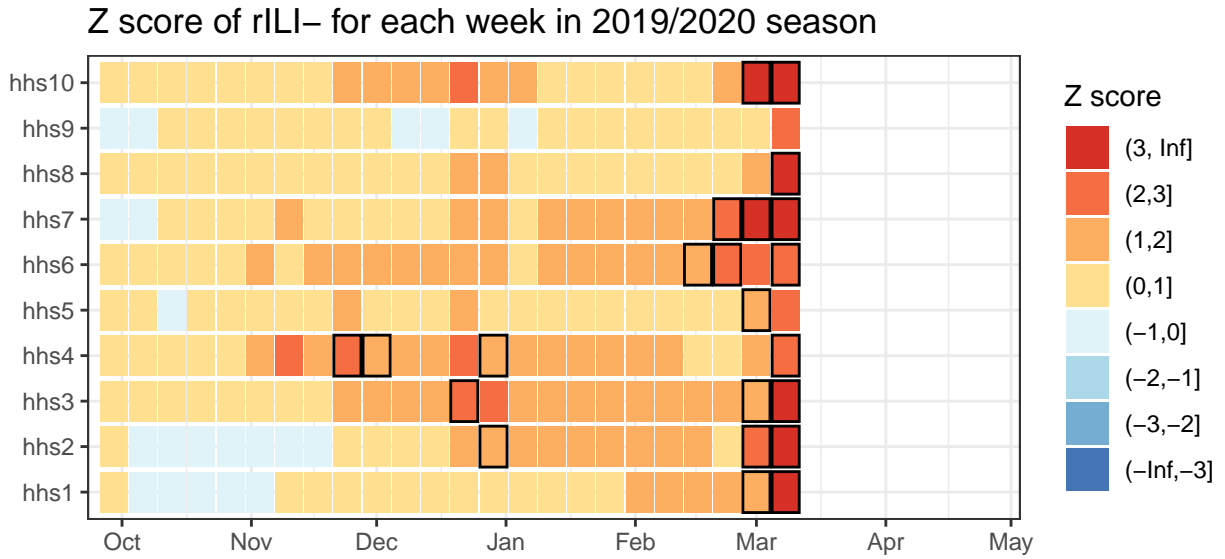
Figure 3: Figure showing Z scores by week for each HHS region. Tiles with a dark black outline indicate locations where the observed rILI- was higher in that week of the season than had ever been observed in the last 22 seasons.

Missouri has recorded relatively few cases of COVID-19 to date. We note that only 9 providers contributed data to the ILI surveillance network in the most recent week of data compared to 11 to 13 in earlier weeks. Provider type (e.g. pediatrician, hospital) is known to influence ILI data. If there are systematic differences in the population seeking care at different providers, that could explain the recently observed spike. We have been in touch with the Missouri Department of Health. They are aware of our finding and have confirmed that they are carefully monitoring for SARS-CoV-2 activity.

In recent weeks, several states have had an uptick in ILI while simultaneously experiencing a decrease in the percentage of influenza tests that are positive (Figure 6). These trends are also suggestive of widespread circulation of an ILI-causing pathogen that is not influenza.

Syndromic surveillance data is best used to guide further investigation, not as a definitive source. Additional study is needed to understand the burden of SARS-CoV-2 across U.S. states. However, we believe that this analysis could be used to guide and target areas for further investigation

Works Cited

[1] https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200319-sitrep-59-covid-19. pdf?sfvrsn=c3dcdef9_2

[2] Yang, Y., Lu, Q., Liu, M., Wang, Y., Zhang, A., Jalali, N., Dean, N., Longini, I., Halloran, M. E., Xu, B., Zhang, X., Wang, L., Liu, W., & Fang, L. (2020). Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. MedRxiv, 2020.02.10.20021675. https://doi.org/10.1101/2020.02.10.20021675

[3] Imai, N., Cori, A., Dorigatti, I., Baguelin, M., Donnelly, C. A., & Riley, S. (n.d.). Report 3: Transmissibility of 2019-nCoV. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/ Imperial-2019-nCoV-transmissibility.pdf.
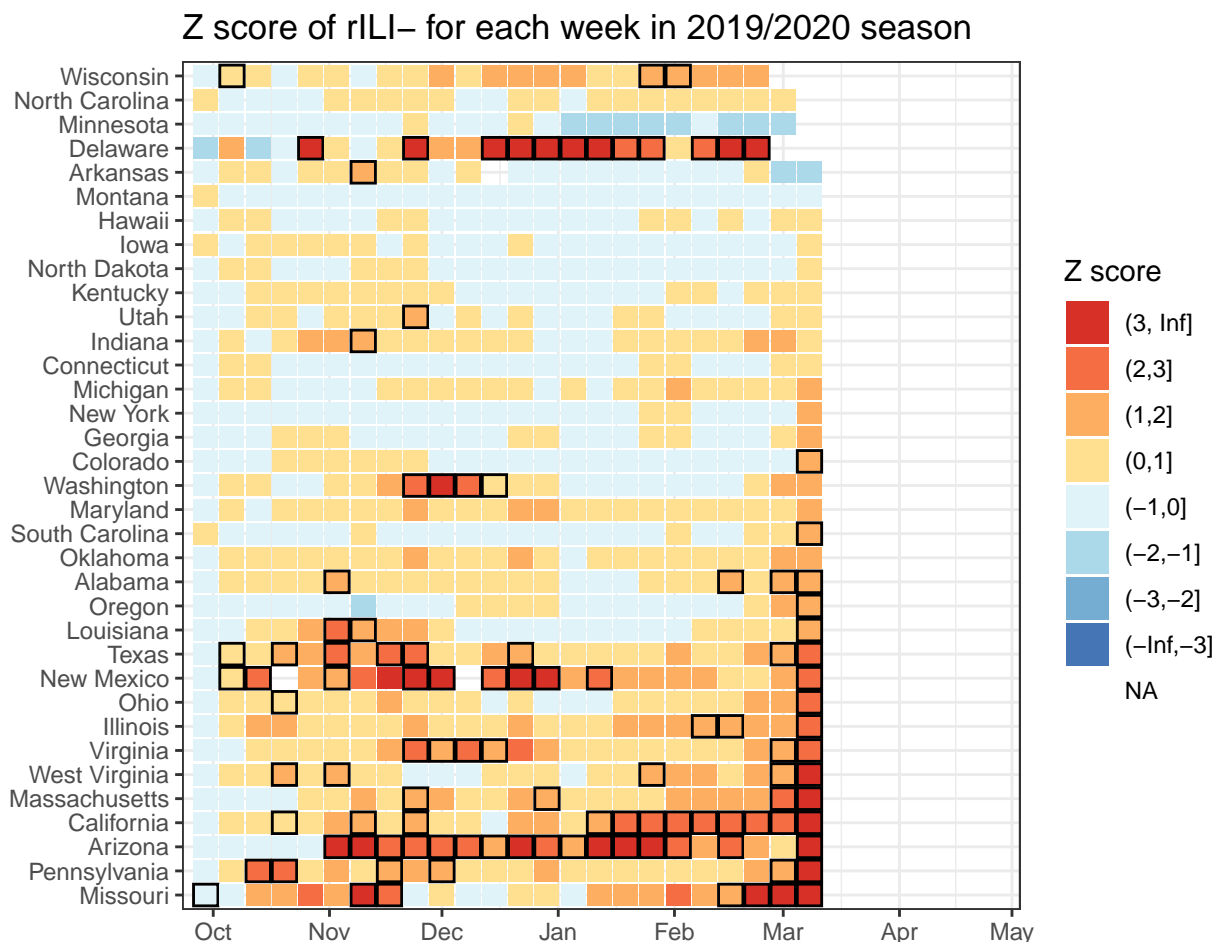
Figure 4: Figure showing Z scores by week for each state in the US that has collected and tested more than 5000 specimens for influenza during the 2019/2020 season. Tiles with a dark black outline indicate locations where the observed rILI- was higher in that week of the season than had ever been observed in the last 22 seasons. Some states have no reported tests for a given week and so the rILI- is missing for that week. States are sorted by z-score in the most recent week, with highest scores at the bottom (states with missing z-scores for the most recent week are at the top).
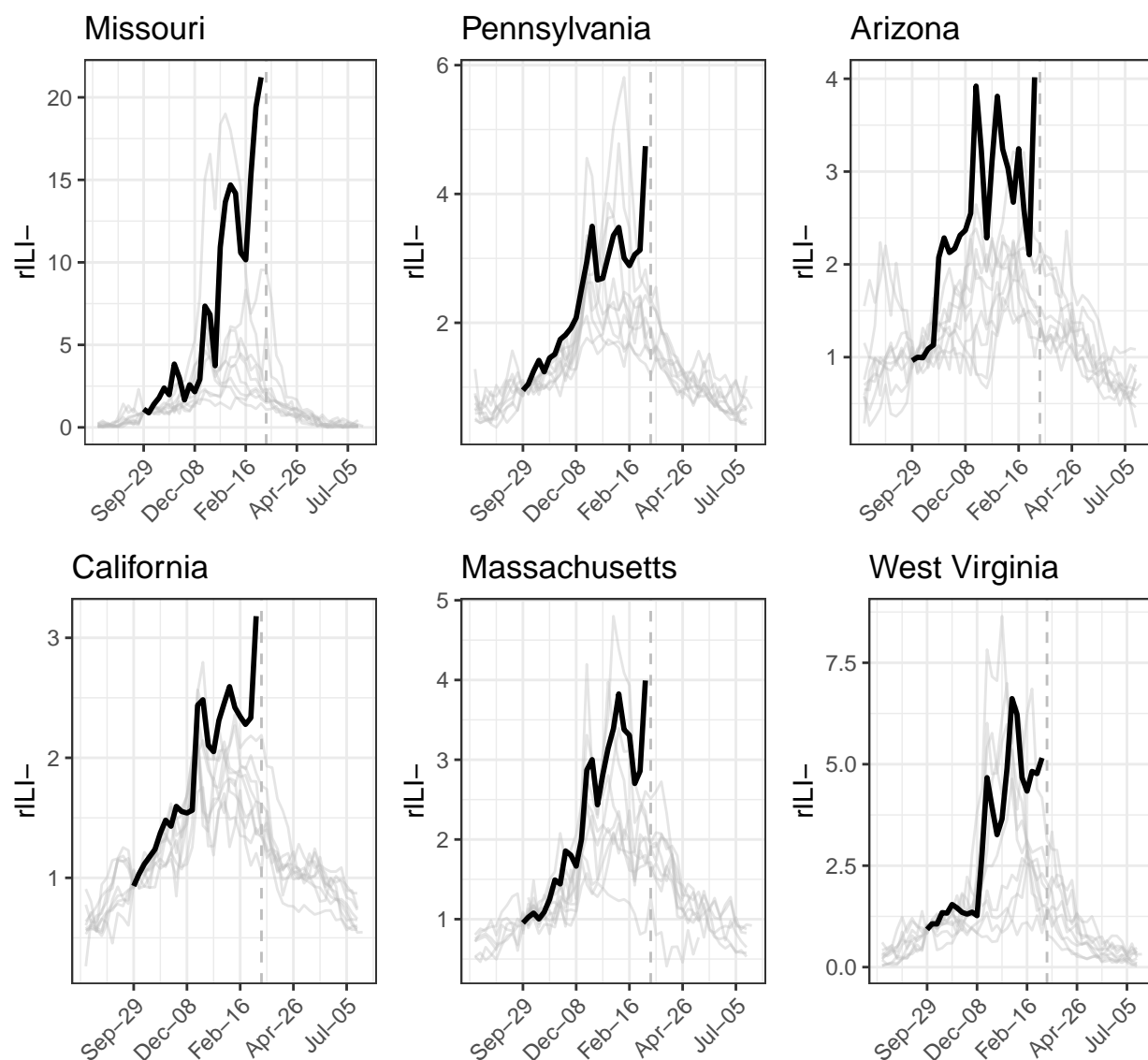
Figure 5: Selected individual state-level plots showing the proportion of ILI not due to influenza over the current 2019/2020 season (dark black line) and past 9 seasons, starting with 2010/2011.
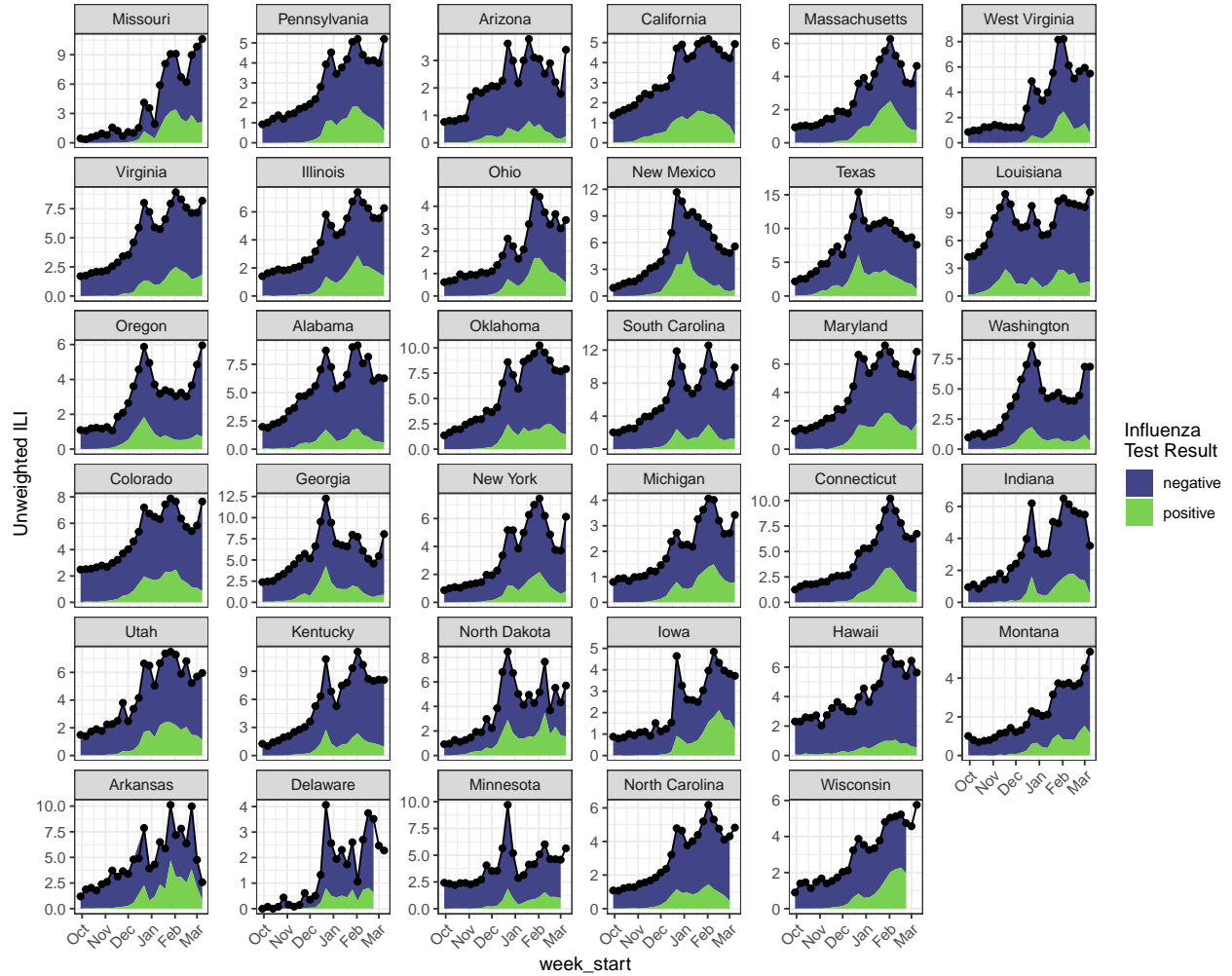
Figure 6: Figure showing inferred breakdown of unweighted ILI into ILI that is positive and negative for influenza for states with greater than 5,000 influenza tests run in the 2019/2020 season.Unweighted ILI positive for influenza is calculated as the product of Unweighted ILI and the proportion of laboratory tests positive for influenza.

Changelog

20 March 2020: Added figure to state level analysis showing estimated breakdown of unweighted ILI into unweighted ILI positive for or negative for influenza. Updated for new ILI data, and reframed the introduction to acknowledge community transmission in the US and discussion to accomodate new results.

13 March 2020: updated for new ILI data, including interpretation about Missouri.

6 March 2020: updated for new ILI data, genomic commentary, some state figures.

2 March 2020: Added state-level analysis, HHS region map, z-score code and figure.

29 February 2020: updated for new ILI data. Minor rephrasing in intro.

21 February 2020: updated for new ILI data.

16 February 2020: updated to revise name of COVID-19, updated case counts and ILINet data, added citations and revised statements about R0.

2 February 2020: Updated to include new ILINet data released on Friday, Jan 31.

26 January 2020: Although our overall assessment has not changed and our analysis has not been updated, we have updated the discussion to better convey the level of uncertainty in our analysis. We also added a heavier line for the 2019/2020 season in the figures.

25 January 2020: First version of report released.