

# 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 Rivers*

*2020-03-13 23:12:10 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 132,758 confirmed cases globally and at least 4,955 deaths from what is currently named COVID-19 [1]. It is now understood that the virus transmits efficiently from person to person, with  $R_0$  estimates above 2 and perhaps as high as 3.7 [2, 3].

Sustained human-to-human transmission has been observed outside of China, and the possibility of unrecognized spread in the United States and other countries cannot be ruled out at this stage. 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. The same analyses suggest that some transmission chains may have been progressing, undetected, in the Pacific northwest of the United States since mid-January.[4] As an early effort to explore the scenario of undetected spread in the United States, 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 circulating unobserved 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, sore throat, etc. . . ). 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:

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

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

## US Health and Human Services Regions

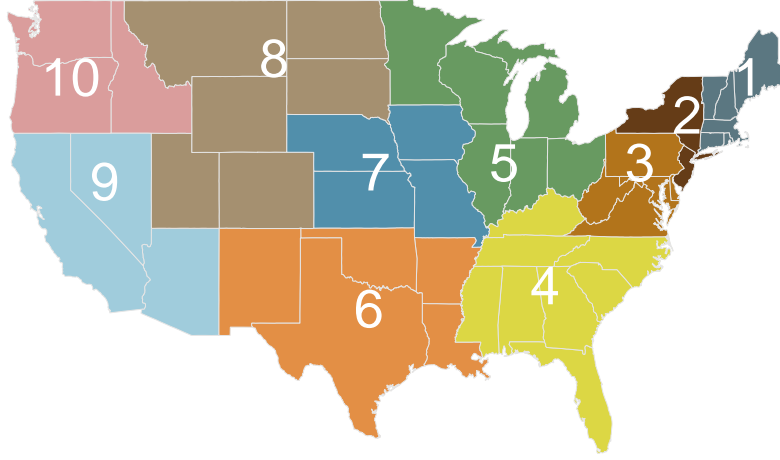


Figure 1: US HHS Regions are made up of groups of states.

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 relative 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:

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

## Results & Discussion

This report uses data downloaded on March 13, 2020, with data reported through March 07, 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. However, the measures of rILI- also have not been consistently more than 3 standard deviations above the average (Figure 3).

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

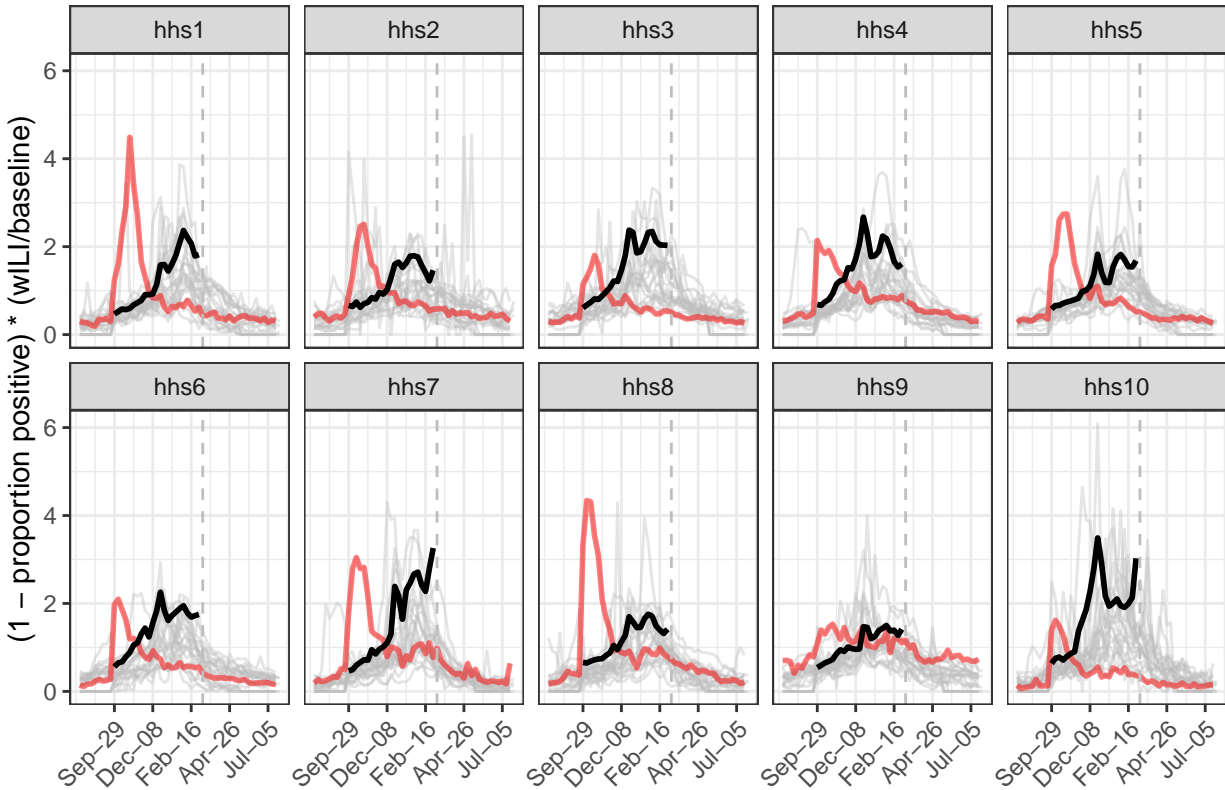


Figure 2: US HHS Regions plots showing rILI- values since the 1997/1998 season (grey lines) and the 2019/2020 season (dark black line). The line highlighted in red is the 2009/2010 H1N1 pandemic season. The dates on the x-axis correspond to the dates for the 2019/2020 season, with previous seasons lined 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.

These results do not particularly rule out any possibilities of COVID-19 transmission occurring in the US at the time of the most recent data reporting or not. If COVID-19 were circulating widely in the US, these data would seem to suggest that its incidence would be currently relatively small, as it would not be adding much relative to levels of rILI- observed in past seasons. 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.

If COVID-19 were to cause significant influenza-like illness in subsequent weeks, we might expect the rILI-metric to increase and be larger than previous seasons. However, 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.

### State-level analyses

The z-score calculations show a few 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.

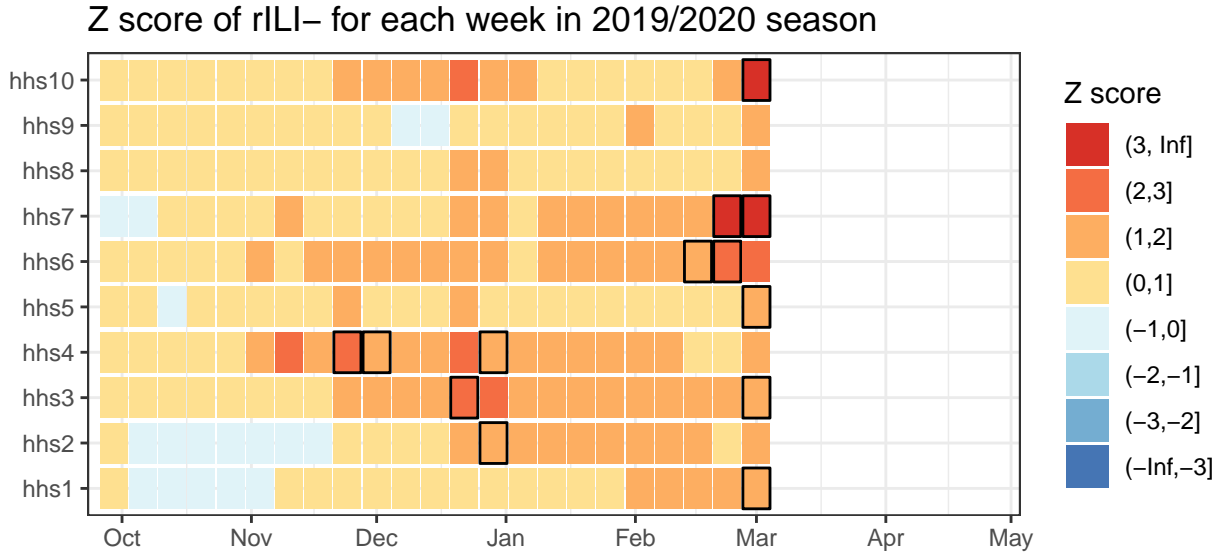


Figure 3: Figure showing Z scores by week for each HHS region. Tiles with a dark black outline indicate locations where the observed rILI- was higher in that week of the season than had ever been observed in the last 22 seasons.

We note an increased signal in California and Missouri (Figure 5). California is known to have community transmission of SARS-CoV-2 (198 cases, 44 thought to be locally acquired as of this writing), but from this analysis alone we are not able to determine whether that is contributing to the observed signal in rILI-.

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.

These results do not necessarily mean that there is circulation of SARS-CoV-2 in these two states. 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.

## Works Cited

- [1] <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
- [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>.
- [4] Bedford T, Neher R, Hadfield J, Hodcroft E, Ilcisin M, M<sup>u</sup>ller N. Genomic analysis of COVID-19 spread. Situation report 2020-03-04. <https://nextstrain.org/narratives/ncov/sit-rep/2020-03-04>

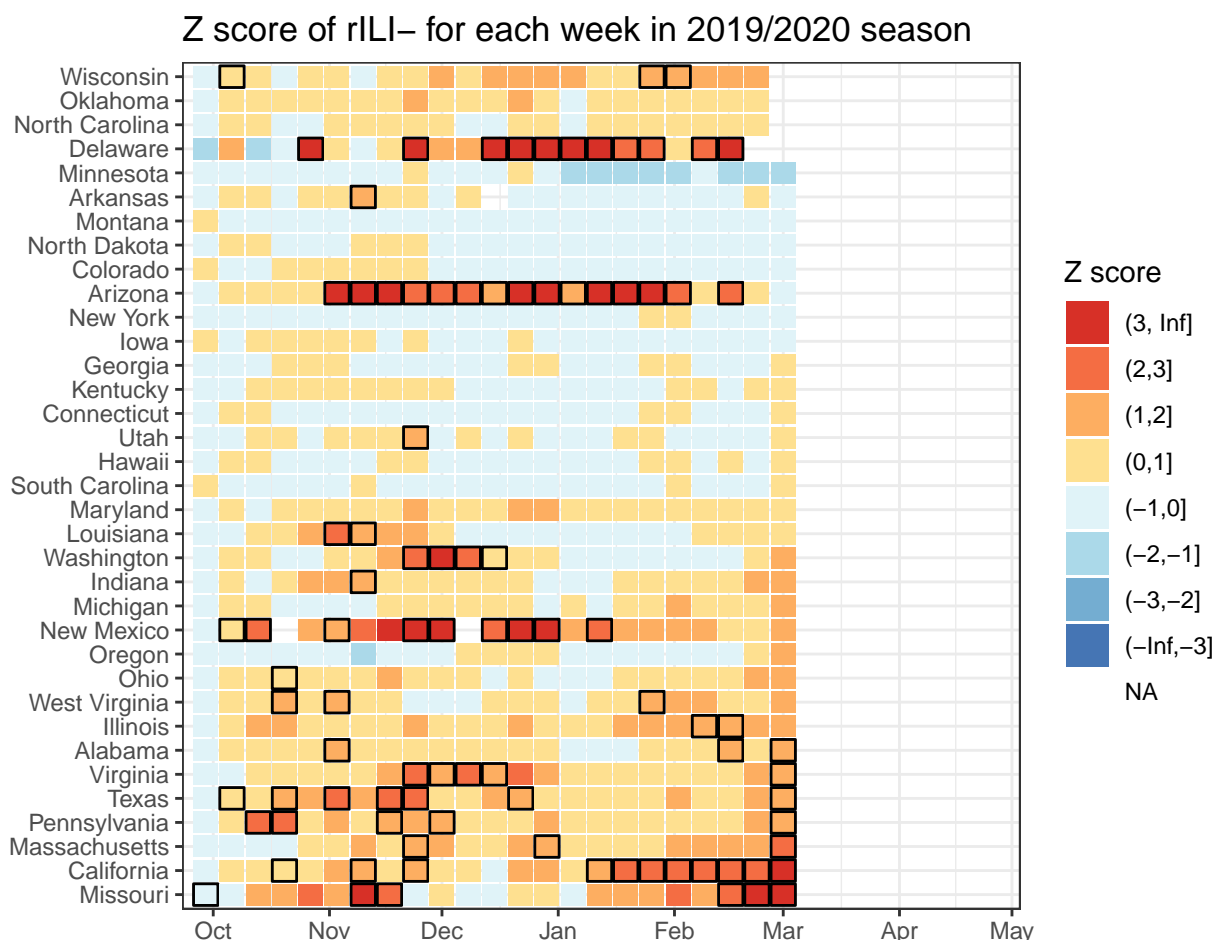


Figure 4: Figure showing Z scores by week for each state in the US that has collected and tested more than 5000 specimens for influenza during the 2019/2020 season. Tiles with a dark black outline indicate locations where the observed rILI- was higher in that week of the season than had ever been observed 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).

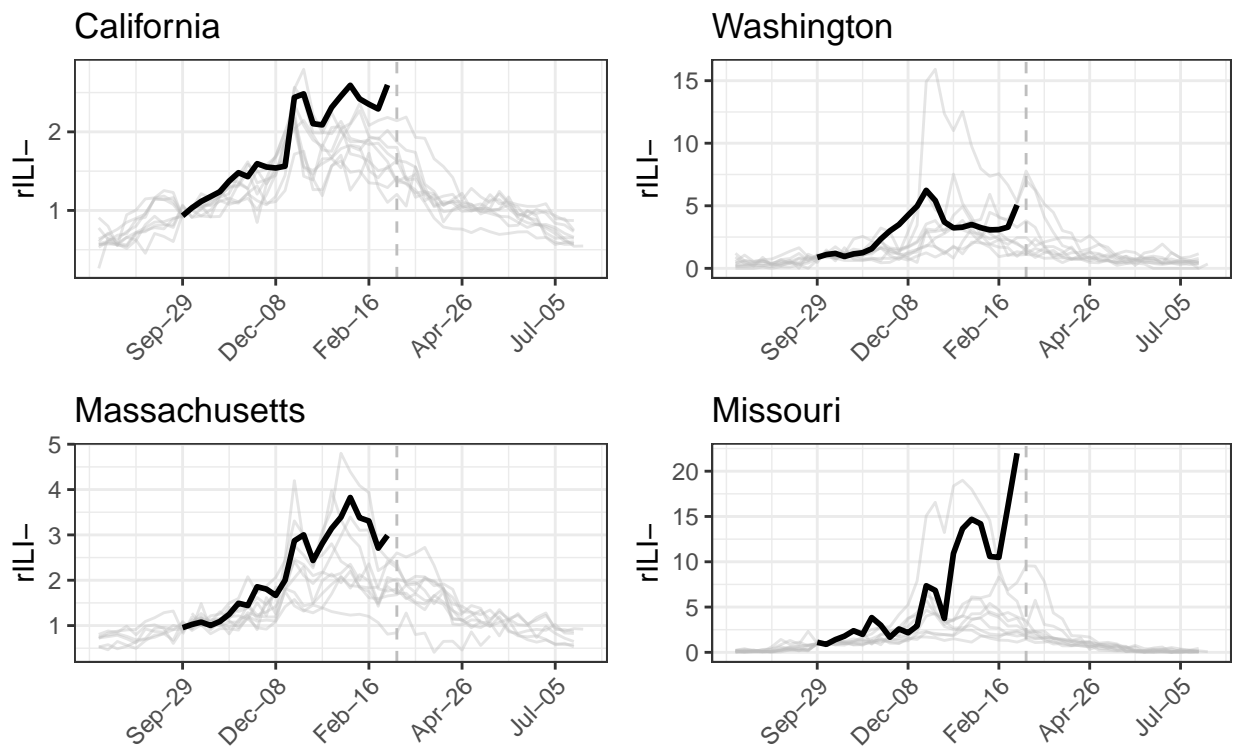


Figure 5: Selected individual state-level plots showing the proportion of ILI not due to influenza over the current 2019/2020 season (dark black line) and past 9 seasons, starting with 2010/2011.

## Changelog

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  $R_0$ .

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.