Normalizing Positivity Rates by Testing Rates *

Michael A. Bailey *Georgetown University*

State-level Covid positivity rates cannot be directly compared due to differences in testing rates across states. This note presents a normalizing method based on a two-stage model of testing.

Keywords: coronavirus prevalence, non-ignorable non-response, survey sampling

It's hard to know what to make of the Covid testing data that emerges from states. For example, look at the figure below that replicates one recently highlighted by Brad DeLong.

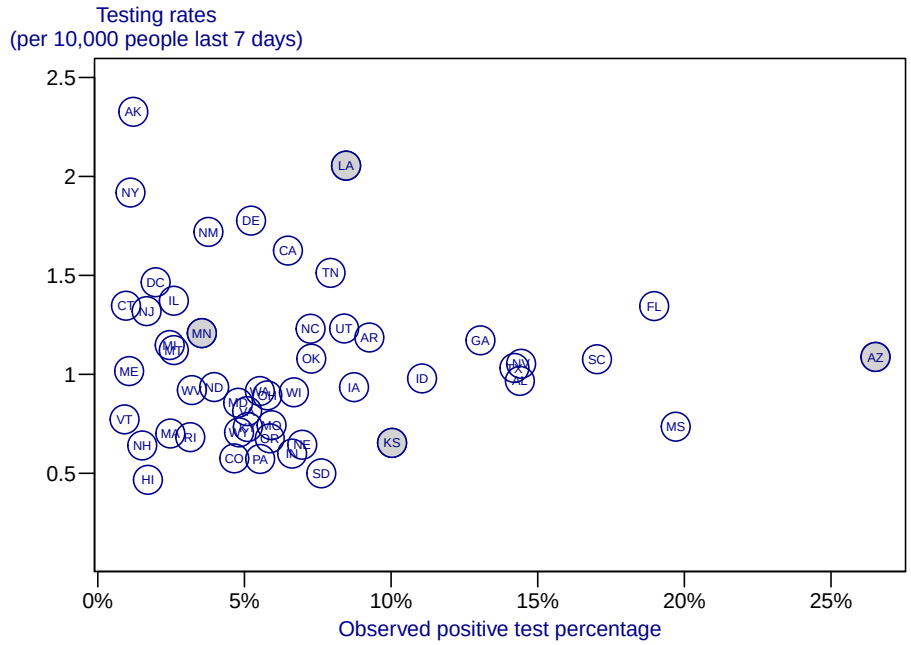

Figure 1: Testing rates and positivity

The positivity rates in Louisiana and Kansas are similar. Should we infer that Covid prevalence is similar as well? Is Arizona really as extreme an outlier as it appears? And can we say anything about a state like Minnesota that is in the middle of a cluster of states with similar testing and positivity rates?

*DRAFT. Please do not cite without permission. Current version: September 06, 2020. Comments welcome to, and code available from Michael.Bailey@georgetown.edu. Errors are my own. Written in RMarkdown building on code provided by Steven V. Miller (http://github.com/svmiller).

These issues arise in many contexts. In India, differences in testing rates across states are even larger than in the U.S. (Khaitan and Bharadwaj, 2020). In addition, there are substantial differences in testing rates and positivity by gender, with men testing at almost twice the rates of women. Among those tested, women have a higher positivity rate. Does that mean that incidence is higher among women? To answer that question, we need to calibrate the observed positivity for the differences in testing rates.

This note presents a method of normalizing positivity rates by the testing rates. The idea is simple: Given observed positivity and testing rates, we can calculate the population prevalence. The population prevalence estimates can then be directly compared or used to estimate the positivity rates expected for any given rate of testing.

This may seem suspiciously simple – and it is, up to a point. It turns out that the mapping of positivity and testing rates to a prevalence estimate requires us to know something very hard to know: the relationship between the probability of being sick and the probability of being tested.

In an ideal world, there is no relationship between the probability of being sick and the probability of being tested. This is the random sampling nirvana hoped for by most researchers (Mostashari and Emanuel, 2020). The reality is that no U.S. state has come anywhere close to implementing such a testing regime. (Iceland has, though (Gudbjartsson et al., 2020) Instead, testing in the U.S. is voluntary, with those more symptomatic or exposed more likely to be tested.

If we know the relationship between the likelihood of being sick and the likelihood of getting tested, it is relatively easy to make testing results comparable across states. I present results here based on a rough, but hopefully reasonable characterization of this relationship. At the end of this note, I present citations on the literature that could help us estimate the key missing ingredient, the relationship between health status and likelihood of getting tested.

In the rest of this note, I present normalized state-level testing results for a plausible value of ρ, the key unknown parameter. I use data from the Covid Tracking Project. Then I discuss caveats; as is often the case with calibration exercises, these estimates should be taken as an improvement rather than as a definitive comparison given the simplifying assumptions necessary. At the end of this note, I formalize the model's logic and discuss some of its limitations.

Normalized test results

The core challenge is that any given state's test results can be consistent with almost any level of prevalence in the state. To see this, consider the case of Arizona in late July, a time when around 1 percent of the people were tested in a week and, of those tested, a remarkable 25 percent tested positive.

How can we translate this information into an estimate of prevalence in the broader Arizona population? If test subjects were chosen randomly, we'd be done: our best guess of prevalence in the state would be 25 percent.

But what if the people who got tested were more likely to be sick than the rest of the popluation? This is almost certainly the case as people tend to get tested when feeling sick or when exposed to other sick people. To understand this case, consider an extreme hypothetical in which every single person in Arizona lined up from most sick to least sick and then the first 1 percent were tested. The first 25 percent of these people tested positive, but after that, only healthy people were tested and they tested negative. And the rest of the untested people are even healthier. In other words, in this case, the prevalence in the state would be estimated to be quite low (roughly the number of people who tested positive divided by the population).

Now consider a diametrically opposite hypothetical in which Arizonians lined up from healthiest to least healthy and the first 1 percent of people in line got tested. Three quarters of these people tested negative, but then the next quarter all tested positive and the entire rest of the state is sicker yet. In this case, the prevalence would be close to 100 percent.

These extreme cases are, of course, unlikely, but they illustrate the challenge. The extent to which health status (the probability of being sick) is related to the tendency to get tested can dramatically influence the way in which we interpret test results. In other words, because the people who get tested are likely unrepresentative of the entire population, we need to make some adjustment to the results in order to generalize to the whole state.

Figure 2 illustrates the issue by presenting prevalence as a function of a parameter we'll call ρ which indicates how strong the relationship between being sick and getting tested is. When ρ equals zero, there is no relationship between health status and getting sick and, as noted above, we can reasonably estimate that 25 of people in Arizona are sick. As ρ gets higher, the population

tested becomes systematically more likley to be sick and we should, therefore, lower our prevalence estimate relative to the $\rho = 0$ case. In the extreme, as noted above, our prevalence estimate becomes very low.

For completeness, Figure 2 also shows estimated prevalence for negative values of ρ, cases in which sick people are *less* likely to get tested. While not impossible, this seems unlikely. But logically, it's useful to see that in this case, the best estimate of prevalence is *higher* than 25 percent.

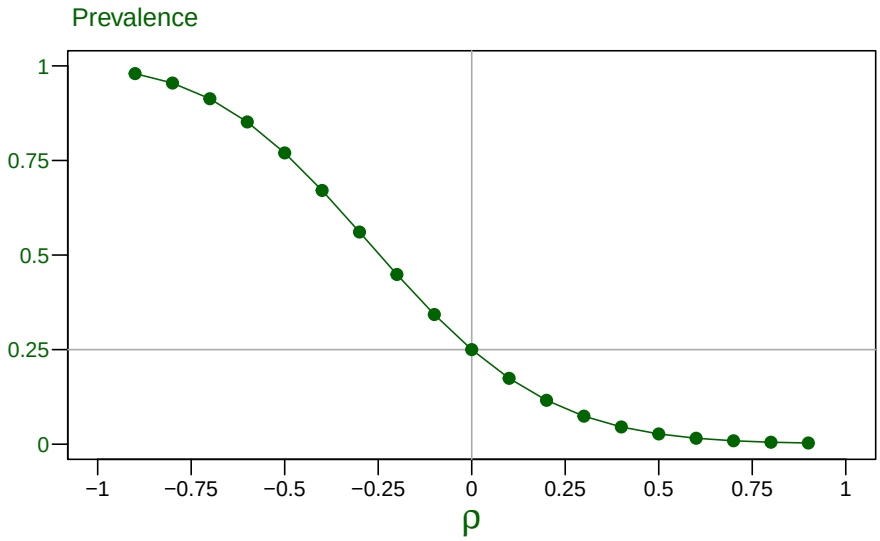

Figure 2: Example: Prevalence as a function of *ho* for Arizona

If we were to know the value of ρ, we can translate test results combined with testing percent into a prevalence estimate. And we can then, in turn, estimate the expected number of cases a state would produce for any given level of testing. In the figure below, we assume a rather strong relationship between the probability of being sick and getting tested ($\rho = 0.75$) and then calculate equivalence lines for various levels of testing. This allows us to, for example, assess what a given state's positivity rate would be for any given level of testing. The positive value of ρ means that positivity rates get higher as testing rates get lower. That is, if we test fewer people and sick people are more likely to get tested, we're likely to have high positive rates when we test only a few people. As a state expands testing, the people getting tested continue to be unrepresentative of the entire

population, but we will nonetheless pull in more healthy people into the testing pool.

To get a sense of how this works, look first at Arizona on the right in red. The red line indicates what the observed positivity rate would be as Arizona changed its testing. Higher testing would lead to a lower positivity rate; less testing would lead to even higher positivity rates. Whatever the level of testing, it is clear that Arizona is in dire straits.

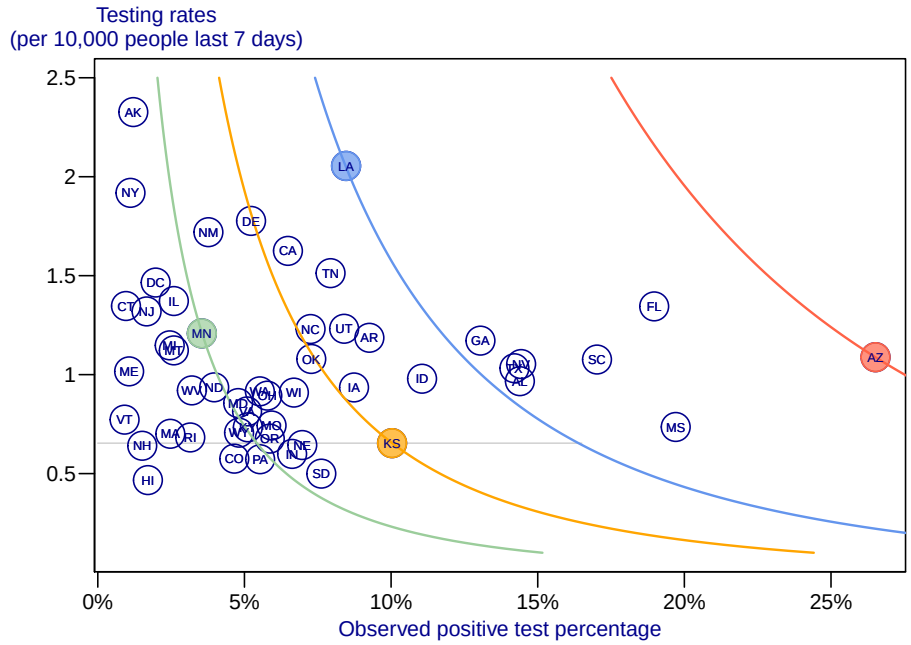

Figure 3: Testing rates and positivity with equivalence lines

Next, look at Louisiana, highlighted in blue Their positivity rate is near that of Kansas, but they test at a higher rate. As we can see, if Louisiana were to test at the same rate as Kansas they would have a clearly higher positivity rate. Specifically, the grey horizontal line in the figure indicates that if Louisiana lowered its testing rate to Kansas's testing rate, Louisiana would have a positivity rate of about 17 percent, which is much higher than Kansas's positivity rate of about 10 percent. In other words, accounting for differences in testing, covid prevalence is higher in Louisiana than in Kansas even though both states have similar positivity rates.

We can also see that once we normalize positivity rates for testing rates, Kansas is quite similar to Oklahoma and Delaware, as these two states are on Kansas's yellow equivalence line.

Finally, look at Minnesota. The state's numbers put it in the middle of cluster of many states. Can we say anything useful? Yes, in some respects. Note that South Dakota is near the green

equivalence line for Minnesota. In other words, despite a substantially higher positivity rate, the low rate of testing in South Dakota means that the prevalence is quite similar as in Minnesota. Colorado has a similar positivity rate as Minnesota, but because it is below Minnesota's equivalence line, covid prevalence there is likely lower there even though Colorado's positivity rate is higher than in Minnesota.

Caveats

This exercise is both constructive and cautionary.

On the constructive side, I show that we can compare positivity rates given a set of not-crazy assumptions. At a minimum, we can use something like this to make positivity rates more comparable than what we currently are able to do.

The exercise is cautionary as well. Comparing positivity rates requires a model and some information that can be hard to come by.

- Most obviously, we need an estimate of ρ, which is the correlation of the error terms in the two equations discussed in the model below. At this point, I have chosen a value of 0.75, reflecting a presumably strong relationship between testing propensity and the probability of being sick. A quick rule of thumb is the higher ρ is, the flatter the equivalence lines becomes. The lower ρ is, the more vertical the lines becomes. In the limit, the lines are vertical when $\rho = 0$, which is the case when sampling is truly random and the positivity rates can be directly compared.

- Hopefully, future work will help us pin down a plausible working value. Estimating ρ empirically is tricky, but not impossible.

- The ρ parameter can vary by state and over time, even within a state. Even as it is plausible that the value is typically within a range that's stable enough for us to have a working estimate, it is possible that this parameter changes. Hence, as testing rates change, it is also possible that ρ changes as well, making it nearly impossible to precisely calibrate positivity rates across time and regions unless we also can measure ρ.

- The model is, as discussed below, based on a bivariate normal functional form. If this functional form is deeply incorrect, the normalizations may be incorrect as well. My intuition is that bivariate normal is almost certainly wrong, but that the normalizations using it will be better than doing nothing in most cases. But this is, of course, an empirical question, requiring further work.
- I have made no adjustments for false positive and false negatives that occur in covid testing.

Model

We begin with a standard two-stage model for testing. The propensity to be tested is

$$R_i^* = \gamma_0 + \gamma_1 X_i + \tau_i$$

where γ_1 is $1xk$ parameter vector, X_i is a $kx1$ vector of covariates and τ_i is a mean-zero random variable. We observe i's test results if $R_i^* > 0$.

The outcome of interest, Y_i, is whether person i has the coronavirus. $Y_i = 1$ if $Y_i^* > 0$ where

$$Y_i^* = \beta_0 + \beta_1 X_i + \epsilon_i$$

The correlation of ϵ_i and τ_i is ρ. Prevalence is a function of the β parameters. For example, if ϵ is normally distributed, estimated prevalence would be the average of $\Phi(\hat{\beta}_0 + \hat{\beta}_1 X_i)$ across the population values of X_i where $\Phi()$ is the CDF of a normal distribution.

We observe $Y_i|_{R_i=1}$, the test results for those who got tested.

Bailey (2020) discusses the literature on selection models of this sort. There are many ways to allow for more flexible functional forms. One general theme in the literature is that the models are unidentified or poorly identified unless there is a first-instrument that affects the probability of testing but does not directly affect the probability of being sick. It is conceptually simple to design a sampling protocol that produces such instruments, although no one has, to my knowledge, undertaken this in the covid context.

Calibrated state level data

The following figures show the case loads, deaths and adjusted case loads for each state. All measures are reported per 100,000 in people. The deaths are in red, reported case loads in blue and adjusted case loads are in green.

The adjusted case loads are calculated in the following manner. First, the above model was used to estimate prevalence for a series of values of ρ. Then, these values were converted to the predicted number of observed cases in the case that the given state tested 2 percent of its population in a given week. This produces adjusted case loads per 100,000 people per week for each state for each value of ρ, the lagged values of which were then used to predict death rates per 100,000 people per week. The value of ρ that produced the highest fit was 0.4. In other words, I selected ρ to maximize the predictive power of the model with regard to future deaths.

The data for most states are scaled from 0 to 400 for the case variables and from 0 to 4 for deaths. However, some states measures exceed these limits for one or both. New York and New Jersey are on a scale of 0 to 900 for cases and 0 to 25 for deaths, for example.

Some data reported by states is clearly anamolous. For example, Colorado reported *negative* 200 deaths on one day in May. Louisiana reported *minus* 2,999 negative results on a single day in April. New Jersey added 1,854 cases in a single day when they started including probable cases. These observations have been excluded and are visible in the plots either gaps in the lines or data that starts at a week later than week 1.

Although these adjusted measures have the clear advantage of increasing the predicted case load in periods of low testing, they should be taken as an improvement on raw data and not as a firm estimate. First, the value of ρ used, while reasonable, is estimated via a calibration that is not particularly precise. It is likely that this value varies across states and time.

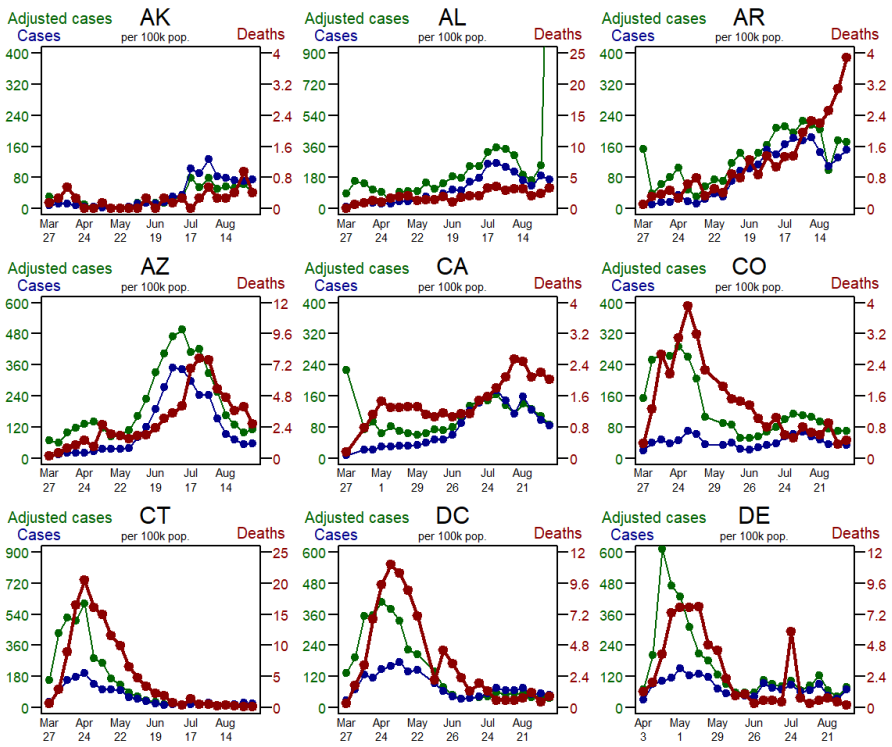

Figure 4: Cases and deaths

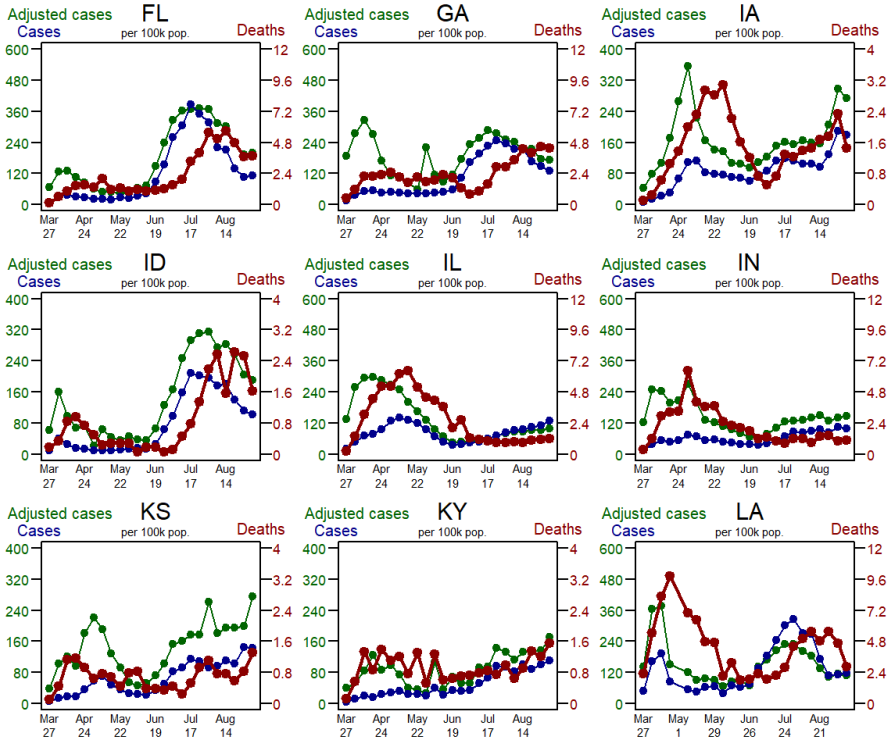

Figure 5: Cases and deaths

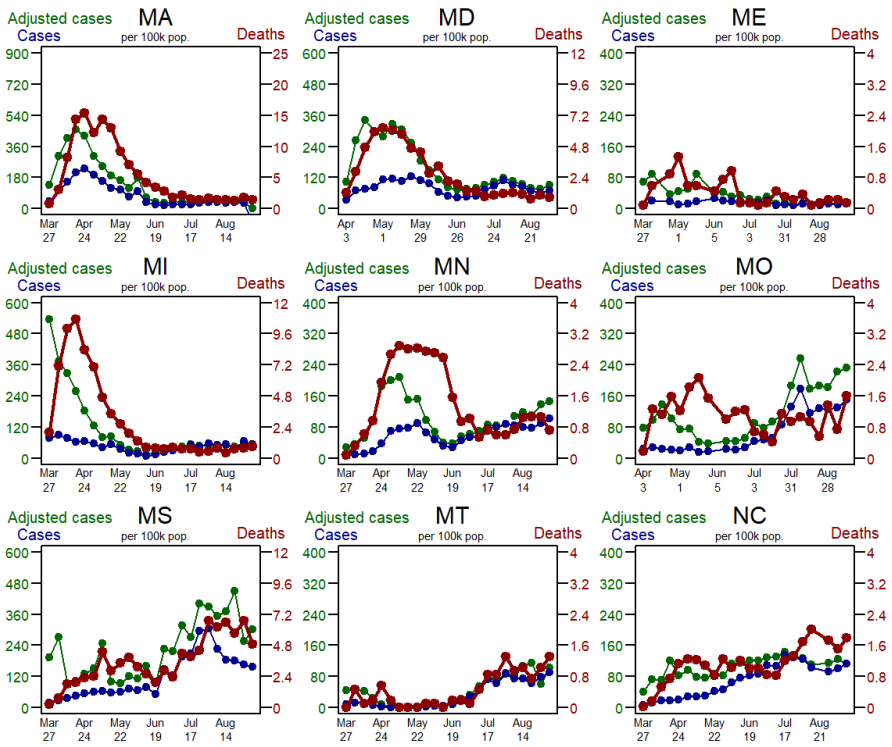

Figure 6: Cases and deaths

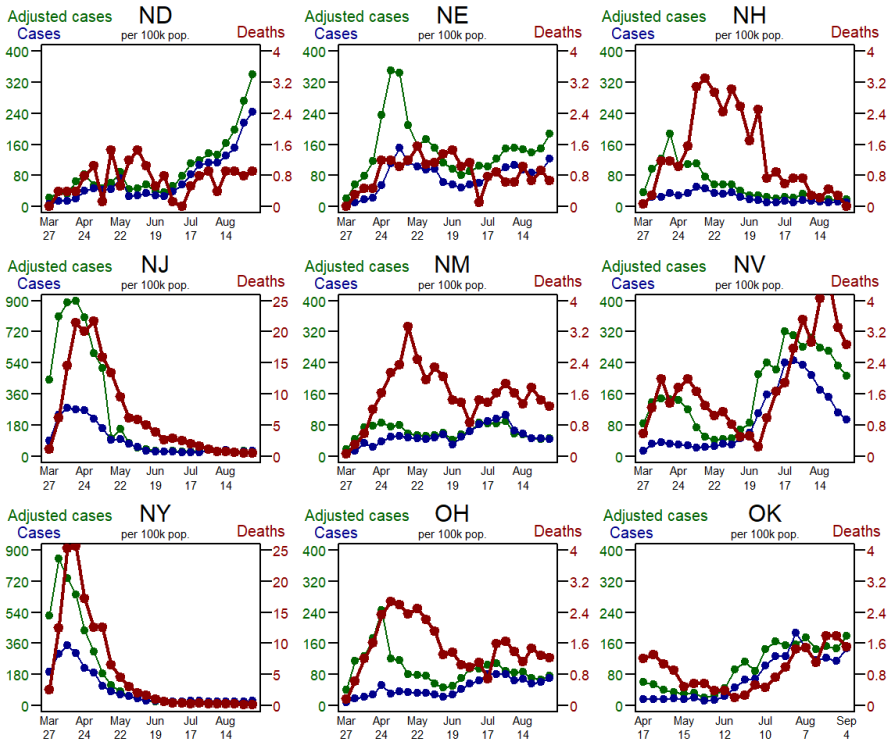

Figure 7: Cases and deaths

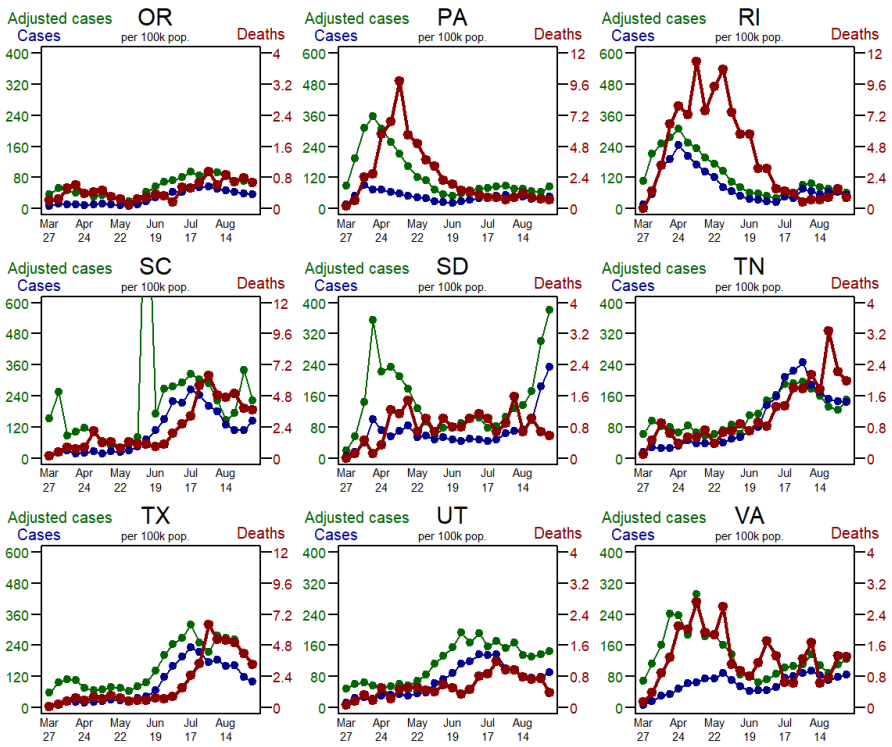

Figure 8: Cases and deaths

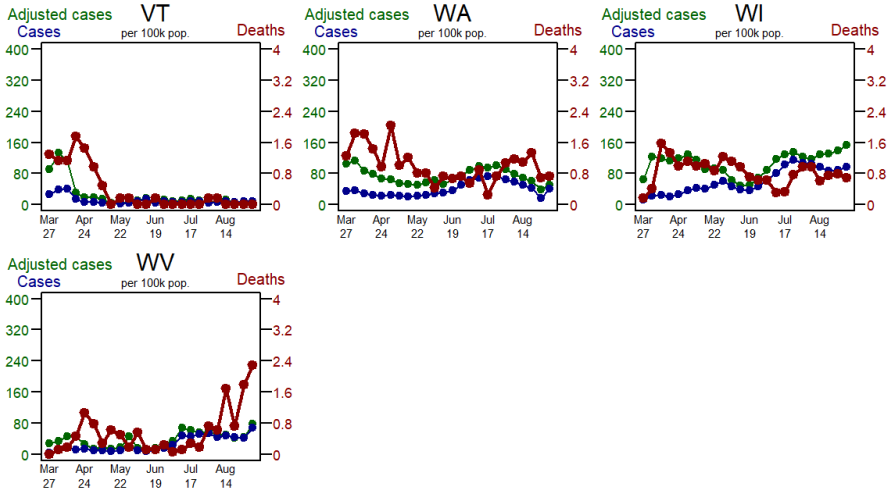

Figure 9: Cases and deaths

References

Gudbjartsson, Daniel F, Agnar Helgason, Hakon Jonsson, Olafur T Magnusson, Pall Melsted, Gudmundur L Norddahl, Jona Saemundsdottir, Asgeir Sigurdsson, Patrick Sulem, Arna B Agustsdottir, Berglind Eiriksdottir, Run Fridriksdottir, Elisabet E Gardarsdottir, Gudmundur Georgsson, Olafia S Gretarsdottir, Kjartan R Gudmundsson, Thora R Gunnarsdottir, Arnaldur Gylfason, Hilma Holm, Brynjar O Jensson, Aslaug Jonasdottir, Frosti Jonsson, Kamilla S Josefsdottir, Thordur Kristjansson, Droplaug N Magnusdottir, Louise le Roux, Gudrun Sigmundsdottir, Gardar Sveinbjornsson, Kristin E Sveinsdottir, Maney Sveinsdottir, Emil A Thorarensen, Bjarni Thorbjornsson, Arthur Love, Gisli Masson, Ingileif Jonsdottir, Alma Moller, Thorolfur Gudnason, Karl G Kristinsson, Unnur Thorsteinsdottir and Kari Stefansson. 2020. "Spread of SARS-CoV-2 in the Icelandic Population." *medRxiv* .
URL: *https://www.medrxiv.org/content/early/2020/03/31/2020.03.26.20044446*

Khaitan, Shreya and Surbhi Bharadwaj. 2020. "Study On COVID-19 Testing Data Calls For Random Testing, Improved Data Quality." *IndiaSpend.com* (June 2).

Mostashari, Farzad and Ezekiel J. Emanuel. 2020. "We Need Smart Coronavirus Testing, Not Just More Testing." *StatNews.com* (March 24).