

Numbers and Uncertainty

Under “Today’s cases” on 5/1/2020 Mike Orsini and Karen Trautmann had a brief exchange about numbers and statistics. They were lamenting the uncertainty of the data and the significance we can draw from the numbers. While they didn’t directly mention it one can infer there is uncertainty in my calculations and graphs. Not only is that correct, in actuality it’s more confusing than they implied.

Why did I write this? Because I always want to be as up front as possible. I’ve tried to make it clear when I’m giving my opinion and when I’m guessing. My goal has always been to give you information, with context, so you can judge for yourself.

One problem is, when people see math, they see it as truth. And in fact, math often is truth but, a common expression in science is garbage in garbage out. Mike Orsini is a scientist (and a good one) so he knows this. His comment about not jumping to conclusions from one number is spot on. But what I’m just learning about is the tenuousness of the data. I’m learning more of it all the time.

So I’ll tell you what I know. Mine you I’m not beating up those who supply the numbers. We are in a war and what we are seeing is the “fog of war”.

Discrepancies

1. County numbers for Franklin vs our/my numbers for Franklin.
 - a. What I have listed in “Today’s cases” and what is on the Twp web site is the total number of Franklin people who lived in Franklin when they tested positive. The county has a different mindset. They are concerned with how many people they are tracking in each town. So, if someone living in Manville tests positive initially, they are counted in Manville. If, however they move to their son’s place in Franklin to convalesce, the Manville number goes down and the Franklin number goes up. What is underlying this is the CAD system. This system keeps track of the whereabouts of those infected. It’s important because if a first responder is asked to go to a house, they need to know if there is some one infected inside. The [map](#) reflects what’s in the CAD. While that number is of interest, I think most people would prefer to keep track the way the town and I do.
 - b. And now a new confounder. Today I was going to update my [comparisons](#) of Franklin to other towns in Somerset Co. As I was collecting the numbers from the [county web site](#) I noticed the numbers went down in some towns, by a lot. While that could be explained by 1a, given I hadn’t done the comparison since April 25th it didn’t seem right that the number of infected people who moved out of a town was greater than the number of new infections. So, I called Dr. Reddy, the County Health Officer. She told me the Attorney General, who oversees first responders, decided cases older than 30 days be removed from the CAD system. The concept makes sense for first responders. Once someone is no longer infectious the first responders need not be forewarned. Thus,

keeping the information in the system is a privacy infringement. The argument is over if 30 days is the correct time period. The CDC suggests the infectious period is 14 days, but the data is sparse, and I wouldn't want to risk my life on it.

- c. And then there is the problem of testing time vs reporting time. We have one patient who was 102 years old. At that age the person is statistically more likely to be female, so I'll assume that. I pick her because only one 102 year old tested positive so I, with reasonable assumptions, can follow her. Her positive test was reported to us on April 21st. I'd been rooting for her. Every day she was not listed amongst those who died my hopes grew. Then on April 30th it was reported she died on April 2nd. So, what was April 21st, the day her result came back? When was she tested? Was it the day she died or some time before? What we really want to know is not the day the report comes back but the day they were tested.
2. Long Term Care Facilities (LTCF) – I think it was Kunal who asked if our LTCF were the cause of the relatively high death rate in Franklin (we are at 8.5%, Somerset Co. 8.6%, NJ 6.1%, USA 5.8% and World 7.1%). The State publishes the LTCF numbers ([excel](#), [pdf](#)). The first thing I noticed was for Franklin some are listed in the town of Franklin and one in the town of Somerset. No big deal, we are used to that. Then I noticed there are only 5 listed and we have 6. To make a long story a little shorter. These numbers are self-reported and it appears each LTCF uses their own criteria. For instance, some are reporting all deaths. Some only COVID deaths. For cases some include only patients, others include staff. Some follow patients if they leave, others do not. Needless to say, I was not able to answer the questions.
3. Increased testing. One reason case numbers may go up is because we are slowly increasing testing. This inherently increases the numbers. We don't have complete reporting on negative test or total tests. If we did, I could adjust for that but unfortunately, we don't. Also, as more tests and easier test are made available then the criteria to get tested will change from only symptomatic to anyone can be tested.

Those are the issues I'm aware of (except ages are missing on five patients). We have reported the issues to those in charge, hopefully they will be cleaned up. Is exact reporting important? Yes. But in defense of those in charge, is it the most important thing? Maybe not. The numbers guide us, but if they are off 5% will that change our decisions when so much is in doubt. The bigger problem is we are still testing very low numbers compared to what we need to be doing. If we are to expend limited energy, I'd rather see it go towards more testing.

Hopefully 1a and b will be fixed, that's local and relatively easy and Dr. Reddy wants to fix it. I will keep analyzing and reporting data. I will correct it as best I can. You just need to understand its limits.