Improving the utility of COVID19 testing data

Christopher M. Hoover, Joshua Schwabb, Maya Peterson, others

## Abstract

COVID19 testing flawed with regards to estimating key epidemiological measures like incidence, prevalence, due to unknown sampling weights among those seeking testing. Testing data can still be used to do XXX, and can be further improved to do YYY by ZZZ.

# Background

### COVID19

Areas across the United States and the world are beginning to reopen following the unprecedented pandemic caused by the emergence of SARS-CoV2. While hospitalizations and daily mortalities due to COVID19 are decreasing in many places, concern remains of a second wave. Widespread availability of testing for active and past infections remains a critical component of the ongoing response to the COVID19 pandemic…

### Testing as passive surveillance

Testing represents a form of passive surveillance commonly used to estimate key epidemiological measures of infection such as disease incidence and prevalence, or rates of transmission such as . However, data generated by such systems is biased. The number of tests available, test-seeking behaviors of different populations/individuals, and [something else] all influence the number of positive tests and the positive test percent in a given location at a given time.

### Goal: Improving the utility of testing

**Problem 1**: Testing is not being used appropriately in many cases, but holds valuable information if analyzed/corrected appropriately, e.g. adjusted for known bias.

**Solution 1**: Propose methods to derive epidemiological information from biased testing data, even as the exact sources and magnitude of bias are unknown.

**Problem 2**: Testing no longer as limited by availability, can testing resources be reallocated/dynamically allocated between passive and active systems to derive better epidemiological estimates/better inform policy?

**Solution 2**: Simulation framework of some sort with:  
1) Data generating process that simulates infection  
2) Testing process that accounts for test availability, variability in test-seeking behaviors, other sources of bias  
3) Estimation of epi measures (incidence, , increasing infections, etc.) from testing data and comparison to truth (from input parameters to data-generating process)

# Correcting for bias

## Incorporating other data sources (hospitalizations, deaths)

* Time-delay limits utility
* See [this LSHTM report](https://cmmid.github.io/topics/covid19/global_cfr_estimates.html) for one method that seems to be widely accepted
* I worked briefly on trying to derive a max likelihood estimate of COVID19 cases on day t from hospitalizations and deaths, but got stuck on right censoring. Could revive that effort though

## Using/fitting models

### Approach Chris has been working on

* Still relies on additional data sources for fitting
* Projections require assumptions of number of tests, bias in testing in the future

# Improving testing

* What additional data can be reported to improve epi estimates (Symptom status, age, etc.)
* Can additional data be collected in order to improve the utility of available testing resources
  + Intuition is that smart reallocation of testing to do some active surveillance will improve estimation even more than just doing more tests, which I think could/should be one of the main messages
  + Could maybe start with data from the Mission Study, use it to derive age/race/income? adjusted rates for all of SF, compare to testing data reported for all of SF county at the same time. Could provide an initial estimate of testing bias. Would be even better if we could do record linkage between Mission Study and other testing data

[1]

# References

1. Lewnard JA, Liu VX, Jackson ML, Schmidt MA, Jewell BL, Flores JP, et al. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in california and washington: Prospective cohort study. BMJ. 2020; m1923. doi:[10.1136/bmj.m1923](https://doi.org/10.1136/bmj.m1923)