Tracking COVID-19 Symptoms and Impact in Real Time: A Survey-Based Surveillance System

Abigail Wozniak[1]

If the coronavirus is to be efficiently contained, it is essential that policymakers have timely data on where it is emerging and how populations are faring under movement restrictions. Epidemiologists forecast that communities will need to actively manage the virus for 18 months or more. Daily data on potential virus spread and community wellbeing are key to ensuring that both containment and social support responses are sufficient.

In recent days, policymakers and researchers have turned to new sources of such data. Often these sources reflect "traces" of virus activity that appear in online applications, like Google searches or personal health apps. This approach is timely, but often incomplete. Statistical agencies and social science researchers have deep experience with safely collecting personal data on health and wellbeing, but it is difficult for them to quickly collect and share data.

This memo outlines how to quickly build a daily survey by combining established population survey methods from social science and public health with advances in electronic data collection in the private sector. By asking about symptoms along with wellbeing, the survey would allow decision-makers to understand in real time: (1) whether and where infection rates might be rising, and (2) how the public is faring under the social and economic restrictions in place.

A successful survey would have the following features:

1. Frequent. The survey must be daily to generate timely information. Because of its frequency, it must also be short.
2. Large. To provide needed detail for state and local policymakers, we need to survey 2 million people daily.
3. Lasting. The full course of the pandemic may be up to two years. Continued daily data will be important in navigating future flare ups and needed restrictions, as well as for gauging the overall recovery.
4. Partnership between existing government survey units and the private sector. Government agencies already have the technical capacity to create samples and questions, and to ensure privacy. The private sector has the capacity to quickly roll out a survey of this size.

5. Three daily modules. Questions should cover key items for decision-makers that are not available with daily frequency elsewhere: mental health and wellbeing, financial security, and physical health.
6. Short and focused, to keep respondent compliance high.
7. Rotating panel. This means a respondent takes the survey one day, is off for a few days, then takes the survey again. Rotating adds important features to the data but reduces respondent fatigue.
8. Recognizes and allows the data quality to evolve.

Details on these features.

Frequent, large, and lasting.

- The epidemic is fast-moving, and restrictions change quickly. Daily data is essential to monitor population health and wellbeing in such a rapidly changing environment.
- Surveying large numbers of individuals is necessary for two reasons. The evolution of the virus, and the impact of restrictions to suppress or contain it, will vary across the country as well as from day to day. Large samples allow for tailored analysis at the state, metro area, and urban/rural levels. To get data that is accurate at these levels of geography, the American Community Survey, our nation's largest survey, interviews over 2 million households annually.
- Large samples may also help address concerns about whether the data can distinguish emerging COVID-19 cases from common symptoms due to other causes. This is discussed more below.
- Large samples may facilitate the use of machine learning to uncover useful patterns in the data in addition to those human analysts set out to study.
- A lasting, 24-month survey plan should allow this information to continue in real-time for the likely duration of the pandemic. Much remains unknown: how will infection rates and severity will progress, what public and economic restrictions will be effective, when can they be relaxed, and how will they affect residents? Daily review of data from this kind of survey could guide real time adjustments of the health and governmental response over the next two years.

Public/Private Partnership.

- National agencies have the expertise to design this survey, but how could it be implemented at large scale? The largest population surveys administered by our statistical agencies require a year to survey their full sample (see table), even though their total sample of half a million to two million respondents is less than 1 percent of the US population.

- In the private sector, by contrast, mobile apps and websites are obviously tracking voluminous amounts of information. Opinion polling and survey technology has also expanded considerably over the last decade. Reaching half a million to two million people daily, and processing their responses, would not have been possible ten years ago. Fortunately, today it is.

- Achieving the needed scale and frequency is outside the abilities of the nation's statistical agencies, but it is well within the reach of multiple technology platforms. For reference, the Pokemon Go app has been downloaded 1 billion times (worldwide). At its peak popularity, its parent company handled daily interactions with the app for over 100 million users. Partnering with a technology platform (or platforms) may also allow for data collection outside the normal scope of surveys. For example, partnering with a health app may allow for even higher quality data collection if the app can monitor body temperature or heart rate.

- Retaining existing agency expertise is essential. Such expertise is critical to including hard-to-reach and underrepresented populations. Partnering with statistical agencies can offer the legal protections for these data that are already part of existing health surveys.

Survey Design Considerations.

- The survey would track COVID-19 symptoms along with indicators of mental health distress and financial hardship. Daily survey questions should fall into three, short modules covering mental and emotional health; financial security, access to services, and impacts of movement restrictions; and physical health. Physical health would primarily ask questions to distinguish emerging COVID-19 cases from other illnesses, but assessment of underlying health issues and questions about known exposure could also be included.

- This approach raises some questions. First, is it possible to track population health by surveying individuals? Yes, in fact many of our nation's measures of population health are collected through surveys. These include rates of various health problems like high blood pressure and heart disease (National Health Interview Survey); illicit drug use (National Survey on Drug Use and Health), and smoking and seatbelt use (Behavioral Risk Factor Surveillance System, BRFSS).

- What about the distinguishing the common symptoms of coronavirus from other ailments? Health providers are already triaging potential coronavirus cases remotely using screener questions. Although COVID-19 shares many symptoms with other common ailments, the CDC identifies the cluster of fever, dry cough, and shortness of breath as the most common COVID-19 symptoms. Daily information on a large sample may help distinguish COVID-19 from other illnesses in the local population, even if it is harder at the individual level.

- A large, daily survey has other advantages for public health research as well. It may help accelerate learning about protective or preventive behaviors. For example, some

commonly prescribed drugs may help combat coronaviruses. A large survey could compare ongoing health of respondents already on those drugs with others. A large daily survey may also help identify clusters of individuals experiencing mild symptoms, like fatigue only, that may indicate COVID-19 spread or re-emergence.

- Other module considerations: Questions on mental health should be asked first, as responses to these are known to be sensitive to preceding questions. Financial and physical health modules could be randomly assigned to be second and third, to address any ordering effects on these topics.

- The mental and financial health modules provide as much critical information as the health module. The daily survey will also provide local policymakers with real time information on how their communities are faring financially and emotionally, including their most vulnerable populations are faring. This will help target support where needed movement restrictions have disproportionate impacts on vulnerable groups.

- A large-scale, rotating panel survey design would represent combine existing government agency expertise with industry scale, and provide many additional advantages. With enough respondents, analysts could measure the distribution of cases in particular geographic areas – states, but also cities versus smaller towns and rural areas. Local policymakers could use this information to tailor their approaches. Rotation balances the additional information of repeated interviewing against respondent fatigue. It would allow analysts to address concerns that regional seasonal variation, like allergies, might falsely increase symptom measurement. And it's possible non-response might be informative or predictive of illness.

- The survey might be structured to interview households rather than individuals. This again involves tradeoffs. A household version would provide greater richness and additional insights into changes in health and wellbeing. However, it adds design complexity and likely makes it harder to build representative data from voluntary respondents, an option discussed next.

- The survey should also be short! A typical response should not take more than 10 minutes although a longer baseline survey might be appropriate. This structure follows the approach in the Current Population Survey. Prepopulating responses with last week's answer after the first wave might also help reduce noise and survey time.

Allow Data Quality to Evolve.

- Ultimately, policy choices should be based on high quality data. But with daily additions to data at a massive scale, data quality can evolve quickly.
- Moving fast and generating public engagement is likely more important at this point than taking time to build a high quality roll out strategy, even if that means initial data waves are of limited usefulness.

- Reweighting, and potentially machine learning techniques, also have the potential to make early, imperfect data more useful. The strategy should be collect now, and adjust, reweight, clean, drop, and mine later.

Finally, will people respond? When many people are scrambling to figure out how to pay their bills, will they take time to answer a survey? Messaging from community and elected leaders is important here. Census Bureau efforts around the 2010 and 2020 Censuses provide a helpful template for how to do this. If Americans are convinced of the importance of this information, my view is they will take time to share it. In doing so, they become part of the solution. To meet significant economic and public health challenges in the past, Americans have switched to smaller cars, planted home gardens, and cut back smoking. History suggests they will take short surveys on their phones.

Survey	Collecting Agency	Sample Size and Design	Design allows sub-state analysis?
Behavioral Risk Factor Surveillance System	CDC, with state PH depts.	Cross section, 400k individuals per year	Yes
National Survey on Drug Use and Health	HHS (SAMHSA)	70,000 individuals	No, not at annual frequency
National Health Interview Survey	Census, CDC partnership	Cross-section, 87,500 individuals in 35,000 households	No
Current Population Survey	BLS and Census	Rotating panel, 60k households per month	Only for 12 largest metro areas
American Community Survey	Census	Cross section, 2+ million housing units per year	Yes

Select References:

https://www.samhsa.gov/data/sites/default/files/NSDUH-MethodSummDefs2014/NSDUH-MethodSummDefs2014.htm

https://www.cdc.gov/nchs/nhis/about_nhis.htm

https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size/

https://www.cdc.gov/brfss/about/brfss_faq.htm

https://healthweather.us

https://journalstar.com/lifestyles/health-med-fit/health/hospitals-pre-screening-patients-to-prevent-spread-of-coronavirus/article_78b6b160-f5a7-5ad5-a015-8972fe11ff2e.html

https://newsroom.clevelandclinic.org/2020/03/19/cleveland-clinic-offers-free-covid-19-online-screening-tool/

https://www.latimes.com/science/story/2020-03-21/coronavirus-treatments-where-we-are-and-what-we-know