GeroScience (2020) 42:1445–1448
https://doi.org/10.1007/s11357-020-00253-6

EDITORIAL

Reflections on designing population surveys for COVID-19 infection and prevalence

Akshay Swaminathan · S. V. Subramanian

Received: 10 August 2020 / Accepted: 12 August 2020 / Published online: 18 August 2020

Introduction

As COVID-19 continue to spread around the world, local and national leaders must closely track three metrics: the total number of active cases, the total number of recovered cases, and the total number of deaths due to COVID-19. These metrics can be used to understand the current risk of infection or mortality and the level of population immunity, and are critical to informing resource allocation and public policy. While many countries have infrastructure in place to record deaths, standard COVID-19 testing procedures that have been used since the beginning of the pandemic cannot reliably capture the total number of active or recovered cases. Since those who exhibit symptoms are more likely to get tested than asymptomatic individuals, confirmed cases are a skewed underestimate of the number of active cases. Furthermore, because standard polymerase chain reaction (PCR) tests for COVID-19 do not detect the presence of antibodies, they cannot detect prior infection. Changes in the availability of COVID-19 testing also affect the number of confirmed cases, which in turn affect the estimated case fatality ratio (CFR); the worldwide CFR of COVID-19 has varied from nearly 10% at the start of April to approximately 2.5% by August (Fig. 1). Without an accurate estimate of the total number of cases in the population, the mortality risk of COVID-19 cannot be accurately measured.

Population-based surveys that test a representative sample of participants using both PCR and antibody tests can be used to estimate both the total number of active cases and recovered cases. The World Health Organization (WHO) recently released a protocol [1] for conducting large-scale serosurveys of COVID-19 for measuring cumulative population immunity and estimating the fraction of asymptomatic, pre-symptomatic or subclinical infections in the population.

In this Editorial, we discuss the contribution of Merkely et al.'s survey of COVID-19 infection rate and prevalence in Hungary [2] in the context of other nationally representative studies of COVID-19, and the key elements of study design that could maximize the value of large-scale COVID-19 surveys for decision-making.

Leveraging institutional collaboration for COVID-19 surveys

Since the outbreak, several countries including the United States (US), Spain, Iceland, Germany, Norway, and India have started or completed national surveys that

A. Swaminathan
Quantitative Sciences, Flatiron Health, New York, NY 10012, USA
e-mail: akshay325@gmail.com

S. V. Subramanian (✉)
Harvard Center for Population and Development Studies, 9 Bow Street, Cambridge, MA 02138, USA
e-mail: svsubram@hsph.harvard.edu

S. V. Subramanian
Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

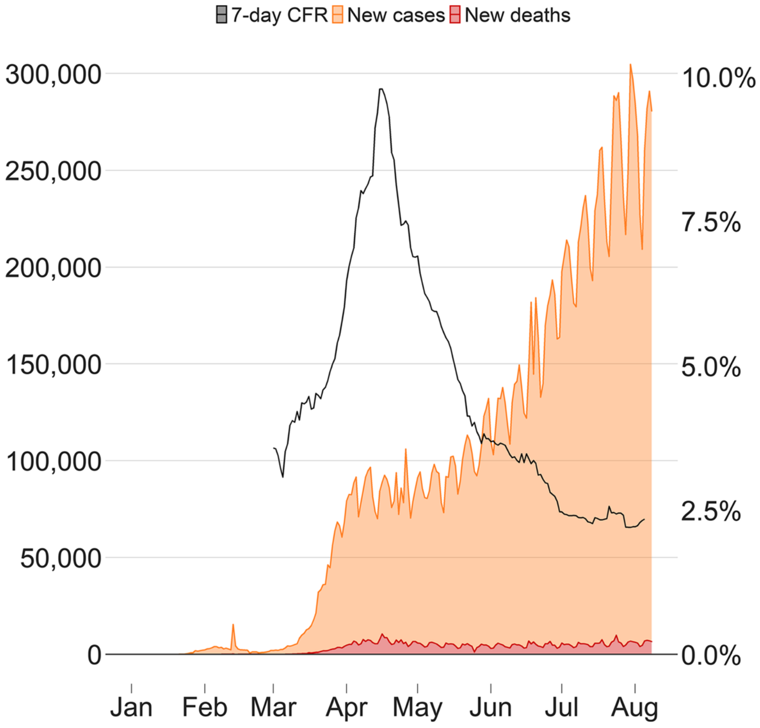

Fig. 1 Worldwide daily new cases and new deaths (both shown on the left *y*-axis), and 7-day case fatality ratio (CFR) (shown on the right *y*-axis) of COVID-19 from January 1, 2020, to August 8, 2020. (Data Source: https://github.com/owid/covid-19-data/tree/master/public/data)

estimate the number of active cases and/or the number of people with antibodies [3]. Merkely et al. report the findings of one such survey conducted in Hungary between May 1 and May 16, 2020, following a 50-day national containment period. Of the over 8 million residents aged 14 or older living in private households, 10,474 participants, randomly selected through a population registry, were tested using PCR and antibody testing. Sampling was stratified by region, and participants were contacted by phone, email, mail, or in-person visit. Of the tested individuals, three had a positive PCR result and 69 had a positive serological result. They concluded that there was a low burden of COVID-19 in Hungary, estimating 2421 active cases of COVID-19 (active infection rate 2.9/10,000) and 56,439 recovered cases (prevalence 68/10,000) [2].

This study is an important contribution to the growing literature on nationally representative surveys of COVID-19 burden. In particular, it is worth highlighting the use of random sampling [4], stratified by region to allow for representative estimates with equal precision across regions. The resources and effort required to conduct a nationally representative survey without relying on sampling infrastructure from existing surveys are substantial. The authors should be commended for coordinating a collaboration between four medical universities, the Hungarian Central Statistical Office, and local municipalities, governmental offices, and ambulance services. Since the survey was conducted following 50 days of mandated quarantine, the results can be used to inform Hungary's approach in relaxing lockdown measures. In future iterations of this survey, the sampling frame could be expanded to individuals not living in private households to increase representativeness, and results could be reported by socioeconomic group to better understand disparities in health outcomes. The sensitivity and specificity of the tests used should also be reported, especially in a country like Hungary where low prevalence of COVID-19 can lead to low positive and negative predictive values without sufficient sensitivity and specificity.

Designing large-scale surveys of COVID-19

As countries prepare to re-open while others prepare for a wave of cases, representative studies that estimate the number of active and recovered cases will continue to play an important role in governmental decision-making. We highlight three elements that are essential in designing informative COVID-19 studies: (1) randomly sampling from an adequate sampling frame; (2) reporting estimates for key demographic subgroups; and (3) conducting repeated measurements across time.

Random sampling and sampling frame

Short of a census, random sampling is the only way to ensure that survey results are representative of the entire population [4]. Since most countries currently prioritize testing of symptomatic individuals, confirmed case counts do not adequately capture asymptomatic or subclinical infections. Disparities in access to testing are another reason why confirmed case counts are not representative of all positive cases [5]. While implementing random sampling at a national level is certainly resource-intensive, many countries can leverage existing infrastructure from previously conducted surveys. Subramanian and James have proposed using the state-of-the-art sampling framework established by the Demographic and Health Surveys (DHS) in the 90+ low- and middle-income countries (LMICs) where DHS have been regularly conducted for decades [4]. Similarly, in a pre-peer-reviewed study, Stringhini et al. report on the results of a serosurvey conducted on 1335 individuals in Geneva, Switzerland [6]. Using existing sampling infrastructure from previous representative surveys, they repeated serotesting weekly for 8 weeks, allowing for continuous tracking of the changing infection rate and prevalence across time.

Equally as important as random sampling is defining an adequate sampling frame. Most countries that attempt nationally representative surveys of COVID-19, including Merkely et al.'s study, exclude institutionalized residents from their sampling. While institutionalized groups may present logistical barriers to data collection, omitting this segment of the population can have especially serious consequences during the COVID-19 pandemic. Residents in prisons and nursing homes are at higher risk for infection and death, and these institutions have been the epicenters of regional outbreaks in the US [7–9]. Furthermore, public awareness of the vulnerability of institutionalized groups to health issues has traditionally been low, so including such groups in COVID-19 surveys can be important in raising awareness and improving conditions.

Considering demographic and socioeconomic heterogeneity

In addition to random sampling across a broad sampling frame, infection rate and prevalence estimates must be reported for key demographic subgroups, such as by gender, race/ethnicity, and socioeconomic status. Subgroup estimates are essential for quantifying disparities in health outcomes that may be due to biological, social, or political reasons. In the absence of random sampling stratified by demographic variables, estimates for demographic subgroups can be calculated by weighting or other post hoc statistical methods. However, analytic adjustment cannot overcome inadequate power arising from insufficient samples in a specific demographic group [10]. Most national surveys on COVID-19 prevalence that employ random sampling stratify by region or geopolitical unit, likely due to logistical convenience. In contrast, a COVID-19 survey in Luxembourg used stratified random sampling by gender and race, but this procedure was likely facilitated by a web application used to group eligible participants by several demographic variables [11].

Need to prepare for repeated measurements

If the goal of representative serosurveys is to paint a realistic picture of the burden of COVID-19, repeated measurements are essential. The highly infectious nature of COVID-19 means that the number of cases can increase dramatically in a matter of days, as exemplified in the US, where over one million new cases were reported in just 2 weeks [12]. Repeated testing should be a fundamental element in the design of any large-scale serosurvey. Many countries have successfully implemented repeated testing of representative samples. In Spain, a nationally representative survey was conducted in three waves, each lasting 2 weeks with a 1-week gap in between [13]. Each participant was tested three times with subsequent samples collected 2–4 weeks after. In Luxembourg, participants will be tested every 2 weeks for the first 5 months, with a final follow-up 1 year after the first test date [11]. In Geneva, participants were tested weekly for 8 weeks [6]. Repeating large-scale testing for COVID-19 can be resource-intensive, especially if both PCR and antibody tests are used. Indeed, leveraging an existing sampling frame as discussed before can be extremely useful in this regard.

Repeating surveys using either PCR or antibody tests can be a cheaper alternative to only conducting the survey once, but deciding which test to repeat requires considering the relative costs and practical value of each. PCR tests may be more expensive to administer because they require specialized machinery and trained operators [14]. On the other hand, repeating a PCR test survey allows for monitoring of active cases, which is more important from a resource allocation perspective than monitoring recovered cases. Repeating antibody testing allows for tracking of the number of total cases over time and can be used to estimate the infection fatality rate of the virus. However, the accuracy of antibody tests can be variable, and it is still unclear for how long COVID-19 antibodies persist after infection [14, 15].

Concluding remarks

In summary, Merkely et al. make a substantial contribution not only to the literature on large-scale studies of COVID-19 prevalence, but also to the Hungarian government's toolkit in measuring and managing COVID-19. We suggest that future studies of COVID-19 in Hungary and elsewhere consider employing random sampling from an inclusive and existing sampling frame, plan surveys that allow robust estimates for demographic and socioeconomic subgroups, and repeat testing to allow for temporal analysis of infection trends. Researchers and policy makers can refer to a website (https://serotracker.com/Data) compiled by Bobrovitz et al. [3] that tracks large-scale COVID-19 serosurveillance projects and summarizes study duration, sampling frame, sampling approach, test details, and risk of bias. As large-scale surveys become more widely used to inform government responses to COVID-19, it is critical that the scientific community align on key principles of study design to best enable precision policy.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. World Health O. Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection, 26 May 2020. Geneva: World Health Organization; 2020 2020. Contract No.: WHO/2019-nCoV/Seroepidemiology/2020.2.
2. Merkely, et al. Geroscience. 2020;42(4):1063–74. https://doi.org/10.1007/s11357-020-00226-9.
3. Bobrovitz N, Arora RK, Yan T, Rahim H, Duarte N, Boucher E, et al. Lessons from a rapid systematic review of early SARS-CoV-2 serosurveys. medRxiv. 2020:2020.05.10.20097451.
4. Subramanian S, James K. Use of the Demographic and Health Survey framework as a population surveillance strategy for COVID-19. Lancet Glob Health. 2020;8:e895.
5. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA. 2020;323(24):2466–7.
6. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al. Repeated seroprevalence of anti-SARS-CoV-2 IgG antibodies in a population-based sample from Geneva, Switzerland. medRxiv. 2020:2020.05.02.20088898.
7. Lopez G. Why US jails and prisons became coronavirus epicenters. VOX [Internet]. 2020 July 20, 2020. Available from: https://www.vox.com/2020/4/22/21228146/coronavirus-pandemic-jails-prisons-epicenters.
8. Debbie Cenziper JJ, Shawn Mulcahy. Nearly 1 in 10 nursing homes nationwide report coronavirus cases. Washington Post [Internet]. 2020 July 20, 2020. Available from: https://www.washingtonpost.com/business/2020/04/20/nearly-one-10-nursing-homes-nationwide-report-coronavirus-outbreaks/.
9. CDC. Preparing for COVID-19 in nursing homes: Centers for Disease Control and Prevention; 2020 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/hcp/long-term-care.html.
10. Little RJ. Post-stratification: a modeler's perspective. J Am Stat Assoc. 1993;88(423):1001–12.
11. Snoeck CJ, Vaillant M, Abdelrahman T, Satagopam VP, Turner JD, Beaumont K, et al. Prevalence of SARS-CoV-2 infection in the Luxembourgish population: the CON-VINCE study. medRxiv. 2020:2020.05.11.20092916.
12. Wamsley L. U.S. hits 4 million cases of coronavirus — adding a million new cases in just 15 days. NPR [Internet]. 2020. Available from: https://www.npr.org/sections/coronavirus-live-updates/2020/07/23/894688178/u-s-hits-4-million-cases-of-coronavirus-adding-a-million-new-cases-in-just-15-da.
13. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. Lancet. 2020.
14. Kubina R, Dziedzic A. Molecular and serological tests for COVID-19. a comparative review of SARS-CoV-2 coronavirus laboratory and point-of-care diagnostics. Diagnostics (Basel). 2020;10(6):434.
15. Jacofsky D, Jacofsky EM, Jacofsky M. Understanding antibody testing for COVID-19. J Arthroplast. 2020;35(7S):S74–81.