Two Pandemics, Two Surveys in the United States and in Spain

Miguel A. Hernán, MD, DrPH, and Raquel Yotti, MD, PhD

ABOUT THE AUTHORS

Miguel A. Hernán is with the departments of epidemiology and biostatistics of the Harvard T. H. Chan School of Public Health, Boston, MA. Raquel Yotti is with the Instituto de Salud Carlos III, Madrid, Spain.

In every pandemic, two important public health questions are asked: how many people have been infected, and how many people have died from the infection? An accurate answer to these questions is surprisingly difficult to obtain for any country.

As described by Morabia in this issue (p. 438), in the fall of 1918, the US Public Health Service started a survey led by Wade Hampton Frost and Edgar Sydenstricker to answer these two questions about the influenza pandemic in the United States. In the spring of 2020, the Spanish Ministry of Health and the departments of health of the 17 Spanish regions started a survey (ENE-COVID), led by the Instituto de Salud Carlos III, to answer these questions about the SARS-CoV-2 pandemic in Spain.[1]

Both epidemiological surveys were carried out in the midst of a pandemic and faced similar logistical challenges. However, the proposed solutions to these challenges varied greatly because the surveys took place in different centuries and within different health systems. A methodological comparison of the national surveys in 1918 United States and 2020 Spain reflects as much the advancement of scientific knowledge as the social improvements of the last 100 years.

HOW TO SELECT A REPRESENTATIVE SAMPLE?

A first challenge for both surveys was how to select a nationally representative sample. The US survey attempted to obtain "a fair sample of the general population" (Morabia quoting Frost, p. 439) by targeting individuals from 18 localities in 82 sections of the country with population ranging from 25 000 to 600 000. A century later, the databases of the National Institute of Statistics were used to randomly select more than 35 000 Spanish households, stratified by province and town size. For ENE-COVID, the progress in data systems made it feasible to select a truly random sample of the population.

HOW TO OBTAIN THE DATA FROM THE SELECTED INDIVIDUALS?

A second challenge concerned the logistics of approaching the selected individuals and recording their information. For the US survey, areas were selected within each locality for house-to-house canvass. Over a four-month period, field personnel interviewed the housewife or other responsible members of each household and ended up collecting information for about 146 000 individuals. For ENE-COVID, the first wave of data collection was completed in two weeks by mobilizing and training 4400 health professionals in more than 1400 primary care centers, as well as creating an information system capable of hosting up to 2000 concurrent users. More than 66 000 individuals (about 75% of those who had previously received an invitation by phone) provided the information to the study personnel at their doctor's office or in their own homes. The interval between the identification of the survey as a national priority and the start of the field work was less than four weeks. ENE-COVID benefitted from 21st-century telecommunications and a distributed health care system with universal coverage, all of which resulted in a high response rate for a population-based survey.

HOW TO DETERMINE WHO WAS INFECTED?

A third challenge was how to define the spread of the virus in the population. The US survey was designed for "ascertaining as accurately as possible the proportion of the population affected" (Morabia quoting Frost, p. 439). By "affected," the 1918 investigators meant the proportion of individuals in the population who had symptomatic disease—that is, those who self-reported having had symptoms of influenza.

Thanks to a century of advances in immunology, ENE-COVID could determine

the proportion of individuals who had developed antibodies against the virus (via either a point-of-care test or a chemiluminescent microparticle immunoassay on serum), which is a proxy for the proportion of infected individuals. The data from this serosurvey were then used to estimate the proportion of both asymptomatic individuals (those who reported no symptoms but had antibodies against the virus) and symptomatic individuals (those with antibodies who self-reported symptoms). The preexistence of a health care system with clinical laboratories around the country, with coordination from the National Centre for Microbiology at the Instituto de Salud Carlos III, allowed rapid transport and analysis of more than 50 000 blood samples.

THE IMPORTANCE OF ASYMPTOMATIC INFECTIONS

The US survey was carried out at a time during which it was not possible to measure serum antibodies, and, thus, the survey data could not directly quantify the spread of the virus in the population. To do so, assumptions are needed about the number of asymptomatic individuals; for example, Morabia assumed that a third of influenza infections were asymptomatic (as estimated in ENE-COVID for SARS-CoV-2).

The impossibility of detecting asymptomatic individuals also has implications for the calculation of mortality. The 1918 US survey data could only be used to estimate the case fatality risk—that is, the proportion of individuals with influenza symptoms who died during the course of the disease.[2] By contrast, the 2020 Spanish serosurvey data could be used to also estimate the infection fatality

risk—that is, the proportion of individuals infected with SARS-CoV-2 (regardless of symptoms) who died.[3] While knowing the case fatality risk is important for clinical purposes, knowing the infection fatality risk in different population groups assists pandemic management: we have no control over who becomes symptomatic after infection, but we can adopt measures to prevent infection.

In summary, Frost and Sydenstricker's retrospective survey was quite impressive given the options at their disposal in 1918. However, a longitudinal serosurvey like ENE-COVID required an additional century of scientific, technological, and social progress. Historical comparisons regarding other aspects of pandemic management—nutritional status of the population, development of diagnostics, therapeutics, and vaccines—lead to the same conclusion: despite the magnitude of human suffering caused by the SARS-CoV-2 pandemic, our generation has been way more fortunate than previous ones. *AJPH*

CORRESPONDENCE

Correspondence should be sent to Miguel A. Hernán, Professor, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

PUBLICATION INFORMATION

CONTRIBUTORS

Both authors contributed equally to this editorial.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to report.

REFERENCES

1. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet*. 2020;396(10250):535–544. https://doi.org/10.1016/S0140-6736(20)31483-5

2. Lipsitch M, Donnelly CA, Fraser C, et al. Potential biases in estimating absolute and relative case-fatality risks during outbreaks. *PLoS Negl Trop Dis*. 2015;9(7):e0003846. https://doi.org/10.1371/journal.pntd.0003846

3. Pastor-Barriuso R, Pérez-Gómez B, Hernán MA, et al.; the ENE-COVID Study Group. Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: nationwide seroepidemiological study. *BMJ*. 2020;371:m4509. https://doi.org/10.1136/bmj.m4509