Invited Commentary | Infectious Diseases

Using Nonprobabilistic Surveys to Measure COVID-19 Incidence

Jacob E. Simmering, PhD

Related article

Author affiliations and article information are listed at the end of this article.

Perhaps the most difficult task in public health is counting cases, and yet it is the most critical task. Public health decision-makers need to know the incidence of a disease as well as the populations and locations most affected in order to allocate resources and mitigate spread. Equally important, they need these data in a timely fashion—knowing what happened weeks ago leaves us ill prepared for managing tomorrow. The need for such information was made clear during the COVID-19 pandemic in the US. Historically, monitoring of infectious diseases such as influenza or COVID-19 has been done by simply counting documented cases. Over the last 2 decades, there has been interest in using other data resources—such as search interest[1,2] or smart thermometer–measured temperatures[3]—to enhance these counts. However, a simpler approach, as described by Santillana and colleagues,[4] is to use nonprobability online surveys. They found a strong correlation between estimates derived from their online survey and both official COVID-19 case counts and wastewater viral concentrations through 2022. After widespread availability, including government-led mass distribution, of rapid tests, they found a considerable decrease in the correlation between their estimates and official counts; however, correlation remained high with wastewater SARS-CoV-2 concentrations. This finding matches the conventional wisdom that at-home testing was associated with undercounting of cases in 2022 and onward.

Wastewater measurement for COVID-19 remains an almost ideal method for surveillance, with functionally no selection bias and minimal time delay[5]; however, wastewater testing has 2 major drawbacks. First, wastewater testing may provide a mean virus concentration over the wastewater system but cannot provide high spatial granularity or demographic characteristics for the cases. Although some organizations were able to conduct building-level wastewater testing (such as in specific dormitory buildings[6]), metropolitan-scale measurement and reporting are the norm. From this information, we cannot tell in what populations or specific locations the virus is circulating.

Second, wastewater surveillance takes time to deploy—tests must be developed, validated, and sent to participating sites. For an endemic pathogen—such as influenza or SARS-CoV-2—this deployment may be relatively easy to implement. However, it will take time to develop a testing framework for newly emerging pathogens, and it may not be practical to monitor these pathogens through wastewater. The early months of the COVID-19 pandemic in the US featured considerable issues with clinical testing—including a recall of testing kits produced by the Centers for Disease Control and Prevention—which may have delayed early detection and response. Support for clinical and wastewater surveillance will occur on a lag for novel pathogens.

On the other hand, an online, nonprobability survey could be easily deployed for a new pathogen while addressing these limitations. Direct survey measurement allows for the easy capture and integration of demographic and other factors, allowing us to identify specific populations where a virus is spreading and causing the most harm. Unlike wastewater surveillance, no testing kit needs to be developed. Adding a question to a general-purpose online survey is easily accomplished. Moreover, even small surveys—such as 1000 people—can provide an accurate and timely estimation of measures of public health concern.[7] As reported in their study, Santillana et al[4] found that the relatively generic nature of their survey reduced error compared with larger, COVID-19–specific efforts. The adaptability of these survey-based approaches was highlighted starting in 2022. Old methods—such as official case counts—became inaccurate and misleading in the new era. However, survey-derived estimates and wastewater viral concentrations remained as correlated in 2022 as before.

Despite the considerable potential of these methods, some challenges remain. As the authors note, the state-level estimates have considerably more data noise than the national estimate. At least some of this seems likely due to decreased sample size and increased sampling noise. The greatest potential of this approach is to understand, and understand quickly, who is affected by a disease. The accuracy of the estimation in demographic subgroups remains to be demonstrated.

Measurement through polls and surveys in an era of extremely low response rates is extremely difficult; however, these data remain incredibly useful. The potential for subgroup analysis by incorporating geography and demographic characteristics will enhance the utility of this information to public health decision-makers. Although much work remains, the efforts by Santillana and colleagues[4] will be critical to creating, validating, and understanding the important role of nonprobability surveys in both monitoring known diseases, such as COVID-19 or influenza, and responding to the next novel pandemic.

ARTICLE INFORMATION

Published: September 30, 2024. doi:10.1001/jamanetworkopen.2024.35407

Corresponding Author: Jacob E. Simmering, PhD, Department of Internal Medicine, University of Iowa, 169 Newton Rd, Iowa City, IA 52240 (jacob-simmering@uiowa.edu).

Author Affiliation: Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City.

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. *Clin Infect Dis*. 2008;47(11):1443-1448. doi:10.1086/593098

2. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012-1014. doi:10.1038/nature07634

3. Miller AC, Peterson RA, Singh I, Pilewski S, Polgreen PM. Improving state-level influenza surveillance by incorporating real-time smartphone-connected thermometer readings across different geographic domains. *Open Forum Infect Diseases*. 2019;6(11):ofz455. doi:10.1093/ofid/ofz455

4. Santillana M, Uslu AA, Urmi T, et al. Tracking COVID-19 infections using survey data on rapid at-home tests. *JAMA Netw Open*. 2024;7(9):e2435442. doi:10.1001/jamanetworkopen.2024.35442

5. Farkas K, Hillary LS, Malham SK, McDonald JE, Jones DL. Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Curr Opin Environ Sci Health*. 2020;17:14-20. doi:10.1016/j.coesh.2020.06.001

6. Godinez A, Hill D, Dandaraw B, et al. High sensitivity and specificity of dormitory-level wastewater surveillance for COVID-19 during fall semester 2020 at Syracuse University, New York. *Int J Environ Res Public Health*. 2022;19(8):4851. doi:10.3390/ijerph19084851

7. Bradley VC, Kuriwaki S, Isakov M, Sejdinovic D, Meng XL, Flaxman S. Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature*. 2021;600(7890):695-700. doi:10.1038/s41586-021-04198-4