International Journal of Epidemiology, 2024, **53**(1), dyad186
https://doi.org/10.1093/ije/dyad186
Opinion

International Epidemiological Association

OXFORD

Opinion

The case for retiring 'period prevalence'

Qiang Xia [1,*] Daniel Bertolino,[1] Muhammad S. Salim,[2] Lucia V Torian[1] and Yingjie Zheng [3]

[1]New York City Department of Health and Mental Hygiene, Bureau of Hepatitis, HIV, and Sexually Transmitted Infections, Queens, NY, USA, [2]Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy, City University of New York, New York, NY, USA and [3]Department of Epidemiology, Fudan University School of Public Health, Shanghai, China

*Corresponding author. New York City Department of Health and Mental Hygiene, Bureau of Hepatitis, HIV, and Sexually Transmitted Infections, 42-09 28th Street, Queens, NY 11101, USA. E-mail: qiangxia@post.harvard.edu

Prevalence, or point prevalence, measures the proportion of people in a population with a given condition at a point in time. The term 'period prevalence' has also appeared in the literature as a measure of the proportion of people with a given condition over an interval of time. However, this term has neither been universally introduced in epidemiology textbooks nor consistently defined in the literature.[1–7] Here, we review the term 'period prevalence' in epidemiology textbooks and online educational materials, and present our arguments against it.

When we measure the prevalence of a condition in an idealized setting, we would conduct a cross-sectional study, measure all study participants at an exact point in time and report the point prevalence. Such a cross-sectional study does not exist in the real world, but we can emulate it by conducting a cross-sectional study over a period of time. In a real-world cross-sectional study, we would obtain point prevalence by treating the study period as a wide 'point in time' or considering all measures as an approximation of the measure at the mid-point of the study. If study participants are evenly recruited, the mid-point of the study is the exact mid-point of the study period on the calendar; if not, it is the weighted mid-point of study participants. For example, when examining the point prevalence of overweight among a group of students, we measure all students once during the study period, treat the study period as a wide 'point in time' and report the point prevalence. We can also consider all such measures an approximation of the measure at the mid-point of the study, and report the point prevalence under the assumption that status changes are negligible or offset by other observations. In our example we can assume that the number of students whose measured overweight status is different from their status at the mid-point of the study is negligible, or that the number of students changing from overweight when they were measured to normal or underweight at the mid-point of the study may offset the number of students where the reverse happened.

Among the epidemiology textbooks that we reviewed, *Gordis Epidemiology*[1] and *Essentials of Epidemiology in Public Health*[2] introduce period prevalence as a prevalence measure, yet *Epidemiology: An Introduction*[3] and *Modern Epidemiology*[4] do not. The World Health Organization (WHO), the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC) include the term in their educational materials.[5–7] Besides period prevalence, the NIH also introduces lifetime prevalence as one of the prevalence measures.[6] Some online educational materials misinterpret incidence proportion as period prevalence. In this paper, we will focus on the two definitions of period prevalence (see Equations 1 and 2) introduced in epidemiology textbooks and by the WHO, the NIH and the CDC.

Equation 1:

$$\textit{Period prevalence} = \frac{\textit{Numbe of existing cases+ Number of new cases during the study period}}{\textit{Number of study participants}}$$

Equation 2:

$$\textit{Period prevalence} = \frac{\textit{Number of people with the condition at any point during a given time period}}{\textit{Number of study participants}}$$

The textbooks and the organizations do not explicitly state the study design to measure period prevalence, but we assume that it should be a cross-sectional study in which all study participants are measured at one time only. In Equation 1, the numerator includes people with the condition when the study starts and people who newly develop the condition during the study period, and the denominator is the number of study participants or mid-interval population. This definition has three issues. First, not all existing cases can be included. Some people can change their status from a case to a non-case after the study starts. For example, an overweight student can lose weight and become a normal weight student on their measurement day. Second, not all new cases can be included. As all study participants are measured once in a cross-sectional study, those who develop the condition after they are measured will not be included in the numerator. Third, the duration of a study affects the measurement. The longer the study, the higher the period prevalence, because there is more time for new cases to develop in time for their measurement day.

Received: 11 July 2023. **Editorial Decision:** 1 December 2023. **Accepted:** 21 December 2023

None of these is an issue in the point prevalence definition outlined in this paper. A prevalence study does not require the inclusion of or distinction between existing and new cases; instead, a study participant's status is simply determined based on one measurement. Including all existing and new cases in the numerator of Equation 1 is not only impractical, but also unnecessary. The duration of a prevalence study in our example has little impact on the prevalence estimate, because only the mid-point matters. For example, a 2-month study from 1 June to 31 July provides an unbiased estimate of the point prevalence at the mid-point (July 1), as does a 12-month study from 1 January to 31 December.

For Equation 2, again in an idealized setting, we would conduct a prevalence study at an exact point in time by measuring all study participants' condition during a given interval of time (e.g. ever diagnosed with SARS-CoV-2 infection in 2021, suicidal thoughts in the past 12 months or lifetime experience of homelessness). Since this ideal prevalence study is conducted at an exact point in time, we would obtain point prevalence. This differs from other studies in that the status of the participants has a time element in it, but we argue that calling it period or lifetime prevalence does not add value but instead causes confusion.

First, this definition obfuscates the three prevalence measures themselves, point, period and lifetime prevalence. Let's consider HIV prevalence as an example. To measure HIV prevalence, we conduct a cross-sectional study, test our study participants and report the point prevalence of HIV infection. Excluding a handful of exceptional cases, HIV sero-status does not change after infection. We can therefore interpret positive HIV test results as ever infected with HIV and negative test results as never infected. Using the period or lifetime prevalence logic and the above interpretation, the point prevalence of HIV infection now becomes lifetime prevalence of HIV infection.

Second, this definition confuses period prevalence with incidence proportion.[4] One reason that researchers measure 'period prevalence' is to quantify the occurrence of a condition that has a short duration, e.g. SARS-CoV-2 infection. A cross-sectional study testing participants for SARS-CoV-2 infection or asking participants if they currently have SARS-CoV-2 infection cannot accurately measure the magnitude of the condition, because of the short duration of SARS-CoV-2 infection. To better quantify its occurrence, we would need a cohort study, in which we follow participants prospectively or retrospectively over an interval of time, e.g. 1 January to 31 December 2021, and report incidence proportion.

Because it is difficult and expensive to conduct a cohort study and sometimes impossible to conduct one retrospectively, some researchers instead conduct a cross-sectional study asking participants if they ever had SARS-CoV-2 infection over an interval of time, e.g. 1 January to 31 December 2021, and report the 'period prevalence' of SARS-CoV-2 infection. As we explained earlier, this 'period prevalence' is actually point prevalence measuring the proportion of study participants at the time of the cross-sectional study who reported having SARS-CoV-2 infection in 2021. Calling the measure period prevalence makes people think it quantifies the occurrence of a condition over an interval of time, but in reality, it only approximates incidence proportion when some assumptions are met, e.g. negligible in- and out-migrations and deaths. If the assumptions are not met, the so-called 'period prevalence' provides limited information about the period, but only the point at which the cross-sectional study is conducted.

It is possible to include both existing and new cases in the numerator of Equation 2, e.g. experience of homelessness in the past 12 months, but the measure from the calculation is still point prevalence, not 'period prevalence'. To avoid confusion, we should call all these measurements prevalence or point prevalence (e.g. the point prevalence of people ever diagnosed with SARS-CoV-2 infection in 2021, the point prevalence of people with suicidal thoughts in the past 12 months and the point prevalence of people with lifetime experience of homelessness).

In conclusion, we reiterate that the prevalence measure from a cross-sectional study is point prevalence. The term 'period prevalence' is not consistently defined in the literature and the introduction of 'period prevalence' does not add value, but causes confusion. We should avoid using the term 'period prevalence' and instead report prevalence or point prevalence from cross-sectional studies.

Data availability

No new data were generated or analysed in support of this research.

Author contributions

Q.X. and Y.Z. conceptualized the paper. Q.X. wrote the first draft. All authors provided critical comments, reviewed the manuscript and approved the final version.

Funding

This work was supported in part by the Centers for Disease Control and Prevention (NU62PS924575 and NU62PS924626).

Acknowledgements

The authors would like to thank Sarah Brunstein, Celia Quinn, Matthew Kuehnert, Tanya Gomez, Patrick Gallahue and Sungwoo Lim for their review and comments on this paper.

Conflict of interest

None declared.

References

1. Celentano D, Szklo M. *Gordis Epidemiology*. 6th edn. Philadelphia, PA: Elsevier, 2018.
2. Aschengrau A, Seage GR. *Essentials of Epidemiology in Public Health*. 4th edn. Burlington, MA: Jones & Bartlett Learning, 2018.
3. Rothman KJ. *Epidemiology: An Introduction*. 2nd edn. New York, NY: Oxford University Press, 2012.
4. Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. *Mordern Epidemiology*. 4th edn. Philadelphia, PA: LWW, 2021.
5. Bonita R, Beaglehole R, Kjellström T. *Basic Epidemiology*. Geneva, Switzerland: World Health Organization, 2006.
6. National Institutes of Mental Health. *What is Prevalence?* https://www.nimh.nih.gov/health/statistics/what-is-prevalence#:~:text=Period%20prevalence%20is%20the%20proportion,is%20a%20commonly%20used%20period (3 March 2023, date last accessed).
7. Dicker RC, Coronado F, Koo D, Parrish RG. *Principles of Epidemiology in Public Health Practice*. 3rd edn. Atlanta, GA: Centers for Disease Control and Prevention, 2012.