Why Confidence Intervals Should be Used in Reporting Studies of Complete Populations

Matthew D. Redelings¹, Frank Sorvillo¹,², Lisa V. Smith*,¹,² and Sander Greenland²,³

¹Los Angeles County Department of Public Health, Office of Health Assessment and Epidemiology, USA
²University of California Los Angeles, Department of Epidemiology, USA
³University of California Los Angeles, Department of Statistics, USA

Abstract: Public-health reports sometimes leave out confidence intervals when data are presented for an entire population. A rationale cited for this practice is that population statistics are measurements rather than estimates; hence there is no need to consider random error because the statistics show exactly what occurred. We argue that this reason does not justify leaving out interval estimates. Targeting intervention in areas with high disease rates can be justified only on the assumption that the excess would continue in those areas; in that case, at the very least, we need to allow for random fluctuations over time. Thus, we recommend that interval estimates be reported even when the entire population is observed.

Keywords: bias, confidence intervals, population studies, random error.

INTERVAL ESTIMATES FOR POPULATIONS

Confidence intervals are used in public-health practice to indicate the degree of uncertainty in estimates due to random error. This makes intuitive sense when estimates are based on data taken from a sample of a larger population, because relations in the sample are unlikely to mirror relations in the true population exactly, even when sampling is random [1]. Some error – random error, at the very least – is likely to occur.

However, official reports sometimes leave out confidence intervals when data are presented for an entire population. This can be justified when case numbers are so large that random error is negligible (e.g., as in large national summaries [2, 3]). With smaller numbers, however, they are sometimes omitted on the grounds that the observed rates are measurements rather than estimates for the population; hence there is no need to consider random error because the statistics show exactly what occurred in the population. [4]

We argue that this reason does not justify leaving out interval estimates. For example, targeting intervention in areas with high disease rates can be justified only on the assumption that the excess would continue in those areas; in that case, at the very least, we need to allow for random fluctuations over time. Those fluctuations can be considerable even when the number of cases seems large, and confidence intervals show the minimum uncertainty needed for applications. We thus recommend interval estimates be reported even if the entire population is observed without selectivity or error. Existence of the latter problems only underscores the need for interval estimation.

WHY CONFIDENCE INTERVALS SHOULD BE A MINIMAL REQUIREMENT

Suppose we knew the exact population size of an administrative region, and the exact number of motor vehicle deaths last year in the region. We could then calculate the exact mortality rate from motor vehicle accidents last year. Nonetheless, when acting as public-health investigators, policy makers, or insurance underwriters, we would not be primarily interested in what happened in last year’s population. Instead, we would want to know what will happen in this year’s population, and the next year’s, because those would contain the events we could influence. We gather population-based statistics retrospectively but use them to make generalizations about current and future situations in populations we believe to be similar. If alcohol-related motor vehicle accidents were an important cause of death last year, we might strengthen drunk-driving laws this year, expecting the trend to continue. If men were much more likely to die in car wrecks than women last year, we will target interventions, such as advertising campaigns, to them in the hopes of preventing deaths in the future.

Even if there is no fundamental change in underlying risk factors in the population, rates and trends are likely to differ somewhat this year from last, as some events randomly occur more often – and others less often – than they did last year. That why we would want to judge how much they are likely to differ before we can plan or set policy. A glance at the confidence limits will indicate a minimal range for reasonable possibilities. Did unlikely events which by chance occurred last year skew our statistics? If so, then even though our observations show exactly what happened last
BIAS AND EXTRAPOLATION ERROR

There are important cautions that are present in all settings, and are not dealt with by confidence intervals. In fact, the term “confidence interval” is misleading in both a narrow technical sense and a broad pragmatic sense. Technically, “confidence” does not refer to the probability that the target parameter of interest is in the interval [1], even though most users seem to interpret it that way (it instead refers to the percentage of times an interval constructed by the same method would contain the parameter across hypothetical unlimited repetitions of an experiment in which the only error is random). But even if we leave this technical distinction aside, we have to confront the fact that our estimates are subject to nonrandom errors. These nonrandom errors can be broadly divided into two types:

1) Internal errors (also known as biases) such as misdiagnosis, underreporting, refusal bias, and confounding, which lead to errors in our estimates even if they are applied only to the population in our study; and

2) Extrapolation (generalization) errors, which arise when the populations or times of interest differ in relevant ways from the population we actually studied.

With either type of nonrandom error, we should not be confident that the true rate or ratio (let alone any extrapolation) falls within the confidence interval, because the interval does not account for errors other than random ones. In particular, when nonrandom errors are likely to be present, confidence intervals do not reflect the uncertainty we should have – thus generating overconfidence in the estimates. In these settings the intervals should at least be accompanied by some warning that they account only for random error (and only for single comparisons at that). Such cautions are most important when the upper and lower confidence limits are close together (i.e., the confidence interval is narrow), for then the confidence interval is most likely to generate overconfident projections.

CONCLUSION

There are many sophisticated methods for accounting for random error in prediction, including shrinkage and machine-learning algorithms [6]. There are also sophisticated methods for expanding and shifting the interval estimate to account for nonrandom errors [7]. Our point here, however, has been to explain why, in the vast majority of settings, at least confidence intervals are needed even when one has captured an entire population in a study. The rationale for surveillance and reporting is to make projections to future events and possibly action on those projections. As a consequence, confidence intervals represent a minimal accounting for the uncertainty that is present even in ideal situations: Uncertainty due to random error.

ACKNOWLEDGEMENTS

None declared.

DISCLAIMERS, CONFLICTS, OR PROPRIETARY INTERESTS

The authors confirm that this article content has no conflicts of interest.
REFERENCES


