

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

Matthew D. Redelings<sup>1</sup>, Frank Sorvillo<sup>1,2</sup>, Lisa V. Smith<sup>\*,1,2</sup> and Sander Greenland<sup>2,3</sup>

<sup>1</sup>Los Angeles County Department of Public Health, Office of Health Assessment and Epidemiology, USA

<sup>2</sup>University of California Los Angeles, Department of Epidemiology, USA

<sup>3</sup>University 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

\*Address correspondence to this author at the Los Angeles County Department of Public Health, Office of Health Assessment and Epidemiology 313 N. Figueroa Street, Suite 127 Los Angeles, CA 90012, U.S.A.; Tel: (213) 240-7785; E-mail: [lsmith@ph.lacounty.gov](mailto:lsmith@ph.lacounty.gov)

year, they will fail to guide us accurately about what will happen this year. Is it realistic to expect that even though last year's mortality rates were higher than the previous year's, the trend could disappear or even be reversed this year? Or is this trend likely to remain largely intact? Similar questions can be asked of observed disease clusters. Confidence intervals help caution answers to such questions.

As an example, published rates of motor vehicle mortality for the District of Columbia (DC) went from 9.2 per 100,000 in 2002 to 10.4 per 100,000 in 2003, a 13% increase. Interpreting such data without recognition of the inherent uncertainties might lead one to infer that there is an upward trend in the rate. In fact, the rate dropped the following year (2004) to 7.6 per 100,000, a 27% decrease, suggesting that the 2003 increase may have reflected only random or other haphazard variation (e.g., adverse weather). In such circumstances, the use of confidence intervals to reflect such nonsystematic variation may help prevent premature conclusions [5].

If we do not indicate the uncertainty in our observations when they are used to make predictions, our efforts to aid research and policy may be compromised. Public health policies are created by generalizing what happened in one population to another supposedly similar population. But we readily recognize that random events in the first population – a lightning strike, a hurricane – may not occur in the second, and vice versa. Though we may observe a population in its entirety, our study is still finite, limited in the number of people it contains and the time during which we observe them. Many events that vary across persons and time, including random ones, will affect our observations.

Random fluctuations may be thought of as the variation due the myriads of unaccounted-for factors that affect individuals but are not systematically related to the aggregate measures we estimate. To account for such fluctuations, we might view a population as a *sample* of a much larger population encompassing other populations and future times of interest. With this perspective, sampling error is still present even when we observe an entire population, since our ultimate goal is to make inferences about these other populations and times. We thus need confidence intervals to give us a sense of how vulnerable our extrapolations are to random variations across these other populations and times.

Even if we do not intend observations from our data to be generalized to other populations, we cannot control how others will use our published data. Thus, when presenting rates and rate comparisons, confidence intervals provide important information for judging the reliability of these observations for planning and policy formation. These intervals can reveal important uncertainty even if the number of events seems large. For example, with 100 events over a year in a region of size 100,000, a rough 95% confidence interval for the rate per 100,000 under a Poisson model<sup>1</sup>, p. 242 would be  $100 \pm 1.96(100)^{1/2} \approx 80$  to 120, spanning a 1.5-fold range.

## 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**

- [1] Rothman KJ, Greenland S, Lash TL. Precision and statistics in epidemiologic studies. Ch. 10. In: Rothman KJ, Greenland S, Lash TL, Eds. *Modern Epidemiology*: 3<sup>rd</sup> ed. Philadelphia: Lippincott-Wolters-Kluwer 2008; pp. 148-67.
- [2] Xu J, Kochanek KD, Murphy SL, Tejada-Vera B. Deaths: Final data for 2007. *Nat Vital Stats Rep* 2010; 58(19): 1-136. Available from: [http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58\\_19.pdf](http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_19.pdf)
- [3] Centers for disease control and prevention. Infant mortality and low birth weight among black and white infants - United States 1980-2000. *MMWR* 2002; 51: 589-92.
- [4] Merrill R. Causal Inference. In: Merrill R, Ed. *Introduction to epidemiology*. Sudbury MA: Jones and Bartlett Learning 2009.
- [5] Motor Vehicle-Related death rates: United States 1999-2005. *Morbidity and mortality. Wkly Rep* 2009; 58(07); pp. 161-5. Retrieved from: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5807a1.htm>
- [6] Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2<sup>nd</sup> ed. New York: Springer 2009.
- [7] Greenland S, Lash TL. Bias analysis. Ch. 19. In: Rothman KJ, Greenland S, Lash TL, Eds. *Modern epidemiology*. 3<sup>rd</sup> ed. Philadelphia: Lippincott-Wolters-Kluwer 2008; pp. 345-80.

---

Received: May 28, 2012

Revised: July 18, 2012

Accepted: July 18, 2012

© Redelings *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.