Head to Head

Response to Engeman: Index values rarely constitute reliable information

By David R. Anderson

In Anderson (2001) I offered an "In My Opinion" paper on "the need to get the basics right in wildlife field studies" and focused attention on the use of convenience sampling and on the use of index values purporting to measure "relative abundance." Engeman (2003) replies that he agrees with the quantitative concepts, but writes to offer a "broader perspective of general statistical rigor, without condemning the use of population indices *if they are appropriately constructed*" (emphasis mine). I thank Engeman for his thoughts on this issue and the chance to make some important points more clearly.

The central logical issue is that an index (commonly an incomplete count) cannot be validly interpreted in terms of some population parameter of interest. Index values are intrinsically unreliable (see Romesburg 1981) or untrustworthy (see Delury 1954) as a basis for valid inference concerning some population parameter. In an important sense, index values are just "numbers," not really "data."

Let the parameter of interest be the size (N) of a well-defined population at a given point in time. An index value is most often a count of some unknown proportion (p) of this population and often not at all representative of the population (e.g., convenience sampling). This unknown proportion is the detection probability ($0<p<1$) and can often vary by an order of magnitude in individual wildlife field studies. (Note: if $p=1$, a census is achieved [the ideal situation].) The detection probability varies for countless reasons and causes (examples in Anderson 2001), and this fact makes it virtually impossible to "appropriately construct" index values, unless p is near 1.

The index value (C), parameter of interest (N), and detection probability are related as $C=N\times p$; solving for the parameter of interest, we have $N=C/p$. If an empirical estimate of the detection probability were available ($\hat{p}$), then the parameter could be estimated simply as $\hat{N}=C/\hat{p}$. It is the empirical estimate of the detection probability that allows the incomplete count (index value) to have meaning and allow a rigorous interpretation.

Over the past century a wide variety of methods have been developed to allow estimates of detection probability as an integral part of study design, field protocol, data collection, and inference (e.g., distance sampling, ratio and regression estimators in classical double sampling, capture-recapture surveys, patch-occupancy models, and sightability models). Use of such methods inherently allow for observer effects, environmental effects, and effects due to characteristics of the target species in the detection probabilities; they are no longer assumption violations. Such detection probabilities are estimated routinely each survey year. Thus, here I certainly agree with Engeman (2003) that survey design and statistical rigor are important, including valid estimates of precision. However, the focus of these efforts ought to be on properly estimating the detection probability each year of the survey. Without empirical estimates of detection probability, index values are not reliable—they are just "numbers." Index values are feckless.

Examples

Consider a simple example in which we have two index values C_1 and C_2, where the subscripts could denote two age, sex, area, or treatment groups and let $C_1=40$ and $C_2=55$. Without knowing or estimating the detection probabilities p_1 and p_2, one can conclude only that $C_1 \leq C_2$ (i.e., that the number 40 is smaller than the number 55) but cannot validly conclude or infer that $N_1 \leq N_2$. It is naive to believe that $N_1 \leq N_2$ based on only the index values $C_1 \leq C_2$. In fact, the opposite may be true ($N_1 \geq N_2$) because $p_1 \leq p_2$ (e.g., 0.35 and 0.55, respectively). Why might group 1 with an index value of 40 actually correspond to the larger population size? Perhaps group 1 is an area with denser vegetation cover and thus relatively few animals are counted (40) because the probability of detection is low, while group 2 is an area with relatively less vegetation and detection probability is higher. Perhaps the observer in group 1 is less capable than the observer in group 2. Unless detection probabilities are known or estimated from the data, index values reveal little about the parameters of interest. No sophisticated analysis procedure can untangle the index as it is a total confounding of the parameter of interest with the unknown detection probability (i.e., N and p are inseparable in N/p).

Commonly, investigators have boldly assumed that detection probabilities were equal across habitat type, observers, years, or species. Such assumptions cannot be checked or verified without additional information (such as estimates of the detection probabilities!). How can one justify *a priori* that the detection probability is the same for differing habitats or differing observers or differing years? Invoking such unsupported assumptions defies both logic and a large literature to the contrary (see Williams et al. 2002: 257–261 for an introduction to these issues). Engeman seems to realize this important point when he states (Engeman 1998: 647), "although making the jump from an index to actual population estimates would require additional measurements to develop 'correction' factors." MacKenzie and Kendall (2002) suggest that the burden of proof should be on showing that detection probability is constant, not the opposite. I wish Hutto and Young (2002) had accepted this burden in their recent paper on perspectives in landbird monitoring.

In many field studies the detection probabilities can be estimated with little additional cost or effort. Notable here are "point counts" used by birders to index population size of various species of birds in the spring (Ralph et al. 1995). These methods provide only index values with the hope that they might reflect, at least roughly, the loose notion of "relative abundance." Such index methods can be replaced if the collection of distance data can be integrated into survey design and field protocol (Rosenstock et al. 2002). Distance sampling (Buckland et al. 2001) allows the estimation of detection probabilities and direct estimation of population size (N) or density (D). These simple changes in field protocol allow the use of the sampling and analysis theory for "point transects." Such methods rely on three assumptions that must be approximately met in the field, and sample size must be nontrivial. Here Engeman (2003) and I agree with White (2001: 383), who wrote, "Don't even start the project if you can't do it right." Why settle for an uninterpretable index value when one can properly estimate the parameter of interest and its precision with little additional resources? In many cases, the estimation of detection probability does not add substantially to the cost of the survey. Of course, exceptions exist where costs and effort may be prohibitive, but I remain skeptical that an index value, while often cheaper, represents any meaningful return.

It is sometimes claimed that long-term monitoring programs can get by with just "indices to relative abundance." Again, this approach is not defensible, for there is often a reasonably high risk that the detection probabilities themselves have time trends. Such trends in detection probability occur, for example, as forests or shrublands grow in height or density over 15–25-year intervals. Roadside habitats have changed markedly over the past 50 years; why should we expect that detection probabilities have remained unchanged as habitats change over such time frames? Other things being equal, if the detection probability tends to decrease over time, the index values will also decrease over time (and vice versa). However, it is entirely possible that the actual population size is decreasing, fairly constant, or increasing. Again, index values do not provide reliable information to allow a valid inference to trends or patterns in population parameters of interest.

Additional clarification

Some investigators ask whether an index can be used if a proper "validation" has been done and the index shown to be related to the actual parameter of interest. Once such a validation has been done, one has the ability to make proper inferences about the parameter using standard ratio or regression estimation theory (Cochran 1963). Thus the index becomes a statistic in an estimator of the parameter. Measures of precision can be easily estimated, and various survey design considerations can be brought to bear to improve future surveys. Such double-sampling procedures should be done annually, or at least frequently, during the course of long-term monitoring.

A *Bulletin* reviewer asked whether sighting distance (a basis for deriving estimates of detection probability in distance sampling) might not change over years in a long-term monitoring program. Indeed, we would expect detection probability (and the related sighting distances) to change substantially over years. If such distances were used to compute an estimate of detection probability ($\hat{p}$) only at the beginning of the monitoring program, then variable detection probability would present a problem. However, in distance sampling, the sighting distances are measured each year as these change in response to a host of year-specific variables. For example, in an open habitat, the average sighting distance might be large and detection probability (assuming a fixed and reasonable transect boundary) might be high (say 0.6 to 0.8). As the habitat changes and substantially more cover appears over time, the average sighting distance will be shorter and detection probability will be lower (say 0.2 to 0.6). Estimates of annual detection probability are continuously updated throughout the multi-year survey program. This is another example of using rigorous methods instead of a crude index.

Another review comment asked about the assumption that animals must be randomly distributed for many estimation approaches. Fortunately, methods such as distance sampling make no assumption about the spatial distribution of animals. Rather, it is the sampling units (square, rectangular, or circular plots, or transect lines or points) that must be placed according to some probabilistic design.

Assumption failures

Engeman (2003) rightfully worries that some methods used to estimate parameters are not robust to the partial violation of certain assumptions. The estimation of population size under the closure assumption is a classic example in which individual heterogeneity in capture probabilities typically causes a substantial negative bias in $\hat{N}$ (Otis et al. 1978). Fortunately, new methods are continually being developed to allow such bias to be reduced (Nichols and Pollock 1983, Nichols 1986). For example, Huggins (1991) proposed the use of individual covariates and conditional likelihoods, and Pledger (2000) developed mixture models as ways to deal with individual heterogeneity in capture probabilities.

Poor survey design

No one condones misuse of methods in research and management. The sampling design and data collection always should be done to a high standard. I find index values to be poor (unacceptable, actually) because of the total confounding of the parameter (N) and the variable detection probability (p). A wide variety of methods now exist to allow estimates of detection probability as part of survey protocol. The routine use of such methods allow the parameter of interest to be estimated directly (e.g., $\hat{N}=C/\hat{p}$) along with a measure of its precision. I believe that the use of index values and convenience sampling are unprofessional and are an enervative approach to important issues in research and management.

Engeman (2003) notes that McKelvey and Pearson (2001) found that 98% of the small-mammal studies reviewed resulted in too little data for valid mark-recapture estimation. This finding, to me, reflects a substantial failure of survey design if these studies were conducted to estimate population size. I feel that their review does not support the use of index values; rather it reflects poor planning and study design in this field (see Hayne 1978 for a similar review). O'Connor (2000) should not wonder "why ecology lags behind biology" when investigators of small-mammal communities commonly (i.e., over 700 cases) achieve sample sizes <10. These are empirical methods; they cannot be expected to perform well without data.

Summary

While Engeman (2003) makes a good case for the importance of study design and that studies that provide only index values are sometimes easier and less expensive, I find I do not agree with his central comments concerning the value of indices. Without estimates of detection probabilities, the use of index values is without a scientific or logical basis. Index values, almost by definition, are not "appropriately constructed"; they are fundamentally flawed. I believe we must focus increased attention on empirical estimates of detection probabilities as an integral part of study design, data collection, estimation, and valid inference.

Acknowledgments. I appreciate comments from P. Lukacs on a draft of this paper.

Literature cited

ANDERSON, D. R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29: 1294–1297.

BUCKLAND, S. T., D. R. ANDERSON, K. P. BURNHAM, J. L. LAAKE, D. L. BORCHERS, AND L. THOMAS. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford, United Kingdom.

COCHRAN, W. G. 1963. Sampling techniques. Second edition.

John Wiley and Sons, Inc., New York, New York, USA.

DeLury, D. R. 1954. On the assumptions underlying estimates of mobile populations. Pages 287-293 *in* Statistics and mathematics in biology. Iowa State College, Ames, USA.

Engeman, R. M. 2003. More on the need to get the basics right: population indices. Wildlife Society Bulletin 31:286-287.

Engeman, R. M., L. Allen, G. O. Zerbe. 1998. Variance estimate for the Allen activity index. Wildlife Research 25:643-648.

Hayne, D. 1978. Experimental designs and statistical analyses. Pages 3-13 *in* D. P. Snyder, editor. Populations of small mammals under natural conditions. Pymatuning Symposium in Ecology, University of Pittsburgh, Vol. 5.

Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47: 725-732.

Hutto, R. L., and J. S. Young. 2002. Regional landbird monitoring: perspectives from the northern Rocky Mountains. Wildlife Society Bulletin 30:738-750.

MacKenzie, D. I., and W. L. Kendall. 2002. How should detection probability be incorporated into estimates of relative abundance? Ecology 83:2387-2393.

McKelvey, K. S., and D. E. Pearson. 2001. Population estimation with sparse data: the role of estimators versus indices revisited. Canadian Journal of Zoology 79:1754-1765.

Nichols, J. D. 1986. On the use of enumeration estimates for interspecific comparisons, with comments on a "trapability" estimator. Journal of Mammalogy 67:590-593.

Nichols, J. D., and K. H. Pollock. 1983. Estimation methodology in contemporary small mammal capture-recapture studies. Journal of Mammalogy 64:253-260.

O'Connor, R. J. 2000. Why ecology lags behind biology. The Scientist 14:35.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62:1-135.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. Biometrics 56:434-442.

Ralph, C. J., J. R. Sauer, and S. Droege. 1995. Monitoring bird populations by point count. United States Department of Agriculture, Forest Service, General Technical Report PSW-149.

Romesburg, H. C. 1981. Wildlife science: gaining reliable knowledge. Journal of Wildlife Management 45:293-313.

Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. 2002. Landbird counting techniques: current practices and an alternative. The Auk 119:46-53.

White, G. C. 2001. Why take calculus? Rigor in wildlife management. Wildlife Society Bulletin 29:380-386.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations: modeling, estimation, and decision making. Academic Press, San Diego, California, USA.

Key words: index values, relative abundance, reliable information, statistical rigor, valid inference

Author's address: Colorado Cooperative Fish and Wildlife Research Unit, Room 201 Wagar Building, Colorado State University, Fort Collins, CO 80523 USA, anderson@cnr.colostate.edu.