What do indicators indicate?1,2

Jean-Pierre Habicht and Rebecca J Stoltzfus

Selection of the best indicators is essential for scientific progress and for implementation of programs. de Pee et al (1), in this issue, show that different investigators rank indicators differently. One explanation is that different procedures reveal different aspects of an indicator. It is important that the aspects revealed are relevant to the indicator's uses. Certain principles are at hand that permit a more systematic approach to reveal these relevant aspects (2, 3).

An indicator is a measurable variable that is supposed to indicate some underlying reality that is more difficult or impossible to measure. One cannot evaluate an indicator without first specifying that underlying reality (4). Serum retinol is an indicator of vitamin A status—the underlying reality. However, in practice our definitions of vitamin A status may vary with the particular question being addressed. Vitamin A status might mean total body stores of vitamin A, vitamin A uptake by one or more target tissues, or vitamin A balance. In public health nutrition the underlying realities of interest are usually not nutrient status per se. When we use indicators to make public health decisions, the correct decision concerns the underlying reality. The decision of interest might concern the number of children at risk of death from vitamin A deficiency and its consequences, or whether a group of women will benefit from vitamin A supplementation, or whether a particular vitamin A intervention had the intended effect. The performance of an indicator will depend on the underlying reality that it is supposed to indicate. This issue is inadequately addressed among nutrition scientists and public health nutritionists.

Once the context of the evaluation is specified, to evaluate an indicator's performance the indicator is compared with a gold standard, which is an approximation of the underlying reality. For an indicator to be useful it must either be summarized for a population by a mean or a rate, or it must be dichotomized by a cutoff value or trigger level, respectively, for classifying individuals or populations. Ranking the performance of indicators requires presentation of the indicators' sensitivity-specificity characteristics over their full range (5) and not only at a single cutoff to be able to test the ranks statistically with adequate power (6). Considerations of these issues (7) and methods to address them (8) can be found elsewhere.

The ideal gold standard for vitamin A indicators is often stated to be liver vitamin A stores (1). In practical situations, this is almost never the case. The fundamental problem lies with the definition of reality, not with its approximation (which is also problematic). The reality must relate to the decision that will be made on the basis of the indicator's measurement. A full review exceeds the space available in this editorial, but a few examples illustrate the approach.

One set of decisions relates to the consequences of vitamin A deficiency. First one needs to decide that vitamin A deficiency is truly an important determinant of the supposed consequences. Then one must decide whether to treat or prevent the consequences through vitamin A interventions, and finally how to monitor the effect of those interventions. An example would be to use the indicators described in the first section of de Pee et al's article to screen mothers for a vitamin A intervention to prevent mortality in their infants. The appropriate gold standard for this decision would be the prevented deaths, and the indicators would be ranked on how well they predicted these prevented deaths. This type of analysis has been done for anthropometric indexes and risk of child death, but not yet for deaths preventable by interventions (2).

A second use of the indicators in the first part of de Pee et al's article might be to screen mothers for maternal interventions to improve infant vitamin A intake from breast milk. The best indicator would best predict the rise in breast-milk vitamin A from the intervention. Thus, the gold standard is the rise in breast-milk vitamin A.

A third use of indicators is to evaluate interventions. This is the subject of the second part of de Pee et al's article. It is important to differentiate indicators of responsiveness to interventions from those indicators that predict benefit from an intervention. The best indicator for these tasks may be different. For example, in an anemic population low serum ferritin may perform well to predict benefit (cure of anemia) from an iron intervention. But ferritin concentrations might not be highly responsive to the intervention because the absorbed iron will be utilized for hemoglobin synthesis, not stored.

The performance of an indicator of response may be influenced by the sort of intervention given, the statistical treatment of the variable, and the range of status of the population. Each of these is illustrated by de Pee et al's study. First, the gold standard used by de Pee et al was the response to a low-dose, five-times-weekly intervention with carotenoids, whereas the

1 From the Division of Nutritional Sciences, Cornell University, Ithaca, NY, and the Center for Human Nutrition, Department of International Health, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, MD.
2 Address reprint requests to RJ Stoltzfus, Center for Human Nutrition, Department of International Health, School of Hygiene and Public Health, The Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205. E-mail: rstoltzf@phnet.sph.jhu.edu.
intervention to which the results were compared (5) was a single high dose of vitamin A (9). This difference may explain the low responsiveness of milk vitamin A in de Pee et al's study. Second, and more important, de Pee et al point out that they used the pre- to postintervention increments in indicator values within individuals, whereas the other analysis used postintervention values. Homeostatically controlled indicators, such as serum retinol, are usually much less variable within an individual than they are across individuals. Thus, they will require much smaller sample sizes to identify differences between groups of women if increments are used rather than absolute values. The opposite is true for nonhomeostatically controlled indicators such as vitamin A in milk. For instance the sample sizes needed to detect differences at 3 mo between postintervention values of groups of women receiving a single high dose of vitamin A or placebo was 148 for serum retinol, 60 for breast-milk vitamin A, and 35 for breast-milk vitamin A as a ratio to fat (5). The respective sample sizes needed for the same comparisons but postintervention increments were 47%, 62%, and 143% as large (RJ Stoltzfus, unpublished observations, 1997).

Third, because of the homeostatic control, serum retinol is less responsive at higher rather than lower levels of vitamin A nutriture. Theoretically, breast-milk vitamin A concentrations will be responsive to vitamin A nutriture at higher concentrations, and thus will tend to respond better to vitamin A interventions than will serum retinol in populations with better vitamin A status. There is some evidence for this in cows (10).

These examples of decisions related to consequences need to be expanded and complemented by examples related to other kinds of decisions, such as those related to nutritional equity or to the search for hypothesized benefits of vitamin A nutriture higher than those necessary to prevent death. For some of these uses liver vitamin A stores might be a useful gold standard, but the case needs to be made that this is so, and if so, that it is practically feasible.

REFERENCES