A problem in health program evaluation is to find a common denominator with respect to which the consequences of programs with diverse objectives and clinical endpoints may be compared. The monetary solution in cost-benefit analysis has slowly given way to the use of the quality-adjusted life year (QALY) as the denominator of a cost-effectiveness (or cost-utility) ratio that can be used to compare the efficiency with which diverse programs use resources to produce health. Implicit in this use of the QALY as the common denominator is the judgment that the QALY is, indeed, a measure of health, as valued from the perspective of the resource-allocating decision maker.

Donaldson et al. (1988) argue that QALY scales developed for a particular context, such as long-term care, are more ‘sensitive’ than scales designed to be applicable to a wide range of programs. They define sensitivity as the ability of the scale to detect statistically meaningful differences between health states. They do not comment, however, on whether these more ‘sensitive’ scales are valid as QALY scales for purposes of program evaluation; that is, whether the QALYs they produce reflect the values of the decision maker (e.g., of the society).

A QALY is a unit on a value function scale (or, for decisions under uncertainty, on a utility scale). Ten years at a quality of 0.5 are judged equivalent to five years at a quality of 1.0. Can this be said of the Life Satisfaction Index? Are ten years with an LSI score of 20 no better than five years with an LSI score of 40? Perhaps some mathematical transformation of the LSI scale would have this property, but previous research has shown that categorical scales not based on explicit time-tradeoff or standard-gamble questions tend to overvalue diminishations in quality relative to death [Torrance (1976)]. The Rosser-Kind scale cited in the article has more face validity as a proper QALY scale for resource-allocation purposes, since it
displays the typical clustering of moderate impairments at the high end (near 1.0), with only the most severe disabilities and symptoms causing losses of utility equal to substantial fractions of the disutility of death.

Donaldson et al. raise several concerns about generalized QALY scales which deserve attention. It is possible that the valuations attached to various losses of quality may differ by age; older people, for example, may value quality relative to life itself more than younger people do. Also, important differences in disease context may require a richer array of possible states than is offered by, say, the $8 \times 4$ Rosser–Kind matrix. These limitations deserve attention in future research.

In summary, then, scales such as the LSI achieve 'sensitivity' by expanding certain ranges of the quality scale to sharpen attention to the areas most pertinent to a particular health care context. In doing so, however, they risk losing validity as proper QALY scales for resource allocation, because they no longer reflect the tradeoffs decision makers would wish to make between the quality and quantity of life. They may not even be valid within the restricted context of resource allocations for long-term care, because the implicit tradeoff between length and quality of life may not reflect either the patients' or society's values. The challenge, then, is to sharpen the ability of generalized QALY scales to reflect age- and context-specific variations in the relative value of health states.

References
