Volume 4 • Number 4 • 2001
VALUE IN HEALTH

Measuring Up to the Task

Paul Kind, MPhil

Outcomes Research Group, Centre for Health Economics, University of York, York, England

The measurement of health status is sometimes perceived as being "soft" and subjective, and is unfavorably compared with the measurement of clinical parameters that are considered as having the desirable attributes of robustness associated with scientific observations obtained under controlled conditions in the laboratory. Such "hard" measurements are reliable, repeatable, accurate, sensitive and above all valid. A positive feast of desiderata! In an effort to achieve parity with its Big Brother, health status measures are often exposed to cruel and unnatural practices that include performing tests that might be regarded as unwarranted under other circumstances. Instrument developers are particularly prone to this tendency since they need to compete in a market that is overly impressed by the size of a *P*-value rather than by any more profound considerations. Hence the tyranny of Cronbach's alpha.

Male sexual health is of increasing relevance and importance, so that new measures for use in evaluating therapeutic interventions in this area ought to be welcomed. One such measure is reported in this edition of Value in Health (Marquis and Marrel). In health care it is both a legal requirement and a moral necessity that products are adequately labeled and appropriately promoted—and not just for the protection of the patient. Health status measures need to be capable of withstanding close scrutiny too. Male sexual (dys)function is simply characterized in terms of four principal elements. Issues not addressed in the paper relate to the generalizability of these elements—specifically, whether the experience of sexual dysfunction in benign prostatic hyperplasia (BPH) is similar to that encountered as a result of other causation. Since the paper lacks any coherent account of the genesis of the MSF-4 (Male Sexual Function 4-item questionnaire), the reader is left to wonder how it is that pre-existing measures are so bad that a new measure is justified. No one would challenge the legitimacy of measuring health-related quality of life (HrQoL) in the treatment of BPH but there has to be a reasonable expectation that instrument developers have a solid case to make before we are asked to junk the old and pick up the new. To do otherwise is a disservice to the appropriate use of HrQoL measurement. Change for change's sake is no rationale for scientists.

For the technically minded, the paper raises long-standing issues that repeatedly figure in HrQoL instruments both old and new. The MSF-4 contains four items with six possible response levels. Each category is scored on a scale from 0 to 5. A global score is computed by aggregating across the four items and this in turn is converted to a score ranging from 0 to 100. The widespread use of such a strategy rests on two basic assumptions. First, that each item has the same value as every other item. Second, that each category within items has the same weight as every other category. Both these assumptions are easily testable. As presented in their native state, they do no more than encapsulate the value judgments of the instrument developer. This, then, is the danger. No invocation of psychometric magic can fully conceal the researcher's influence. So why should we worry about Cronbach when there are bigger issues at stake? The principal characteristic and advantage of HrQoL measurement is that it provides opportunity for patient values and preferences to be taken into account. This advantage is discarded when we revert to the paternalism of the instrument developer. Worse still, we compromise the status of HrQoL measurement by providing our critics with yet one more soft target.

And the lessons to be learned from this? That in the as yet unregulated market for HrQoL measures, there is room for some discipline and self-restraint. Before developing new measures we should carefully study the evidence regarding the performance of the current technology. We must be able to convince potential end users on the basis of conclusive evidence. If it ain't broke don't fix it! "New" does not automatically mean improved. And finally, the question of labeling. HrQoL measures developed in clinical trials suffer

Address correspondence to: Paul Kind, MPhil, Principal Investigator, Outcomes Research Group, Centre for Health Economics, University of York, York YO10 5DD, England. Email: pk1@york.ac.uk

 1098-3015/01/$15.00/279 279–280

from limitations with respect to being protocol driven. Furthermore, there is always the possibility that a skeptical audience might unfairly see an element of self-interest if a new measure indicates positive advantage for an experimental treatment under study. If we are to gain the upper hand in our struggle to achieve the respect of our scientific peers, then HrQoL measures need more than the tired rehearsal of questionable statistics—we need demonstrable rigor in our methods.