

Selecting Performance Measures by Consensus: An Appropriate Extension of the Delphi Method?

Audrey Burnam, Ph.D.

In this issue of *Psychiatric Services*, Addington and colleagues (1) describe an innovative method for choosing performance measures for early psychosis treatment services, the use of a Delphi group consensus method.

The Delphi group consensus method has a long history of development and use by researchers at RAND and the University of California, Los Angeles (UCLA) to measure the quality of health care (2). In the typical RAND/UCLA application of the method, a panel consisting of expert medical research practitioners is asked to rate the extent to which a particular health intervention for a defined group of patients is appropriate on the basis of empirical evidence and practitioners' clinical experience, in which "appropriate" means that the expected benefits of the health intervention outweigh the harms and "inappropriate" means that expected harms outweigh benefits. Only when a high degree of consensus among experts is found for appropriate ratings are these practices used to define measures of quality of care or health care performance.

Because the expert panelists are generally nationally recognized researchers as well as practicing treatment providers and because they are given a comprehensive summary of the relevant scientific literature as part of the rating task, their ratings tend to rely heavily on the efficacy and effectiveness literature. The resulting measures of the quality of care are developed only for areas of practice in which the

evidence base is relatively strong.

The method used by Addington and colleagues is a departure from the RAND/UCLA appropriateness method in major respects. First, panelists were not given summaries of the clinical efficacy and effectiveness literature. The authors note that the early psychosis treatment model has not yet been demonstrated to have clear-cut benefits relative to treatment as usual. Second, the group of raters was asked to rate the importance of each of a list of performance measures for early psychosis treatment services. Third, the panelists were selected to represent seven different stakeholder groups. The method used thus identifies performance measures that a diverse set of stakeholders agree are important.

On what basis were these panelists rating importance of a performance measure in the absence of an evidence base for evaluating the benefits and harms of particular practice components of early psychosis treatment programs? We cannot know, but I suspect that it was some combination of personal preferences, notions about what works best, and an implicit assumption that having a wider array of services is preferable to having few available services.

Is this a good way to go about establishing performance measures for a promising but undemonstrated treatment approach? For clinical processes of care, I think not. Measures should be based on evidence that shows that providing the indicated care (for example, case management) positively affects health outcomes. Expert panelists have a potential role in this kind of task, but the task would involve cautiously generalizing from the literature

on the basis of clinical experience and translating the evidence into useful performance measures—along the lines of the UCLA/RAND appropriateness method. When the evidence is weak, no amount of opinion will help.

On the other hand, the research by Addington and colleagues included performance indicators that represented a variety of outcomes, including various dimensions of patient functioning and quality of life, satisfaction with care, and costs. For outcomes of care, a selection method that values diverse stakeholder preferences and incorporates their priorities into the definition of the bottom-line product of treatment services makes a great deal of sense. I applaud the authors for their efforts and progress in this regard. Too often clinical services and programs are evaluated only on the basis of what matters most to physicians (symptom reduction) or payers (costs) rather than what matters most to patients and families (functioning and quality of life).

The distinction between processes and outcomes of care is an important one, and one that Addington and colleagues failed to make. Broad social and stakeholder values should guide us in establishing the outcomes that are important to measure for mental health care. But once key outcomes are defined, we must turn to evidence to learn how best to produce outcomes that we care about. ♦

References

1. Addington D, McKenzie E, Addington J, et al: Performance measures for early psychosis treatment services. *Psychiatric Services* 56:1570–1582, 2005
2. Fitch K, Bernstein SJ, Aguilar MD, et al: The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, RAND, 2001

Dr. Burnam is a senior behavioral scientist at RAND, 1776 Main Street, P.O. Box 2138, Santa Monica, California 90407-2138 (e-mail, audrey_burnam@rand.org).