



The Importance of PROs in Evaluation of New Medicines

Robert J. Temple, MD
Deputy Center Director for Clinical Science
FDA/Center for Drug Evaluation and Research

Presented at:

FIRST ANNUAL

PATIENT-REPORTED OUTCOMES (PRO) CONSORTIUM WORKSHOP

March 23, 2010 – Bethesda, MD

Co-sponsored by



Why PRO's Now?



Interest in PROs is plainly growing. In part this has to do with:

- 1) Growing interest in “real world” studies and real world measurements.**
- 2) Growing interest in the “value” of treatments, and what better measure of value is there than what the patients themselves notice.**

But there are some other important things to like about PROs, and they are good reasons to use PROs more

- 1) They could greatly increase trial efficiency because they are less dependent on investigator skills.**
- 2) They have arisen in a time of far greater attention to the symptom scales we use.**

As a Real World Measurement of Treatment Value



Consider angina (not much interest lately, but there used to be). Approvals were based on:

- 1. An exercise test (usually using the Bruce Protocol), a graded test in which failure was rapid after a slope or speed increase and an effective agent produced an increase of perhaps 20 seconds, obviously not easily translated to real life.**
- 2. Angina rate – a sensible measure but a problem because many people would exercise until they got angina, so even if they could exercise more, attack rate was little changed.**

As a Real World Measurement of Treatment Value (cont)



In those days there was no attempt to try to ask about how angina affected ADL or anything else and one gained little sense of clinical impact, even if the studies did show clear evidence of “increased exercise tolerance” or, in some case, increased maximal oxygen utilization, both plausible measures of the desired effect.

Similar approaches were used in CHF and assessment of peripheral artery disease, and asthma medications were assessed with pulmonary function tests, again a good measure of an effect, but not easily translated into symptom improvement. Not surprisingly, CHF and asthma assessments were early in the development of PROs, so that these demonstrated benefits could be better translated into symptom and functional improvement.

Efficiency



The idea that, at least in some cases, a skilled physician could “translate” patient reports into a scalar measure or a global” assessment may not be unreasonable but it is critically dependent on the physician’s skill, not only as a doctor but as an interviewer, and it is very easy to imagine that such skill would be variable, especially in a world of growing study size and increasingly diverse locations.

If this gives “noisier” measurements, and greater variability, study power is inevitably decreased. It seems probable (not proved, to my knowledge), that a well-developed PRO, evaluated for consistency and not dependent on external expertise, will show far less of such variability, an attractive feature, especially if trials become more “real world.”

Development Process



Whether because they're recent or because clinicians are more worried about patient-derived data than clinician derived data, PROs receive attention to what they measure, the consistency of measures, and the meaning of differences of a certain size that in many cases far outstrips what we know about more traditional scales, perhaps excepting some of the psychiatric scales that do seem to have a history of very careful development (although "physician globals" were often used too).

This seems particularly true for the disease-specific scales (CHF, asthma, arthritis), probably most pertinent to drug development where extensive efforts have been made to determine what is really critical to patients' lives and to the state of their disease.

Development Process (cont)



The FDA's PRO guidance, written with the experience of many years, describes an assessment of content validity and reliability that is very demanding in order to assure that the test measures what it should (i.e., content validity), is applicable in most environments, and all in all, can be successfully used to evaluate treatments. There seems little doubt that physician based assessments deserve similar treatment.

This seems less clear for many physician-determined scales. NYHA classification of CHF, for example, seems completely reasonable, but I wonder whether, in the areas we're interested in (people in high II and III) it distinguishes effects as well as PRO scales do.

Safety Evaluation



A recent NEJM paper (Basch) suggested that adverse effects too could benefit from more direct patient reporting. Whether that is generally true seem undecided as yet, but it seems likely that such complaints as sedation/sleepiness, fatigue and other subjective complaints, at least once identified, as problems, could benefit from such approaches to better determine their severity and impact. We know that assessment of sexual function demands special attention to determine its characteristics and even frequency.

Future



- 1. Like any clinical measure, the effect of the mean is our typical primary endpoint, but the distribution of effects is of equal or greater interest. This deserves more attention.**
- 2. Although PROs can have multiple domains, to date there has not been great attention to the specific areas that drive the result, or whether this might differ between treatments. This is, however, of interest, just as it is on a HamD (where we so far have never let anyone claim specific effects), and may deserve more attention.**
- 3. It would be of great interest in studies using both physician and patient measures to see whether, as I hypothesized the PROs are really more efficient, less variable, etc.**