Interpretation of Quality-of-Life Scores

Mary-Margaret Chren

Patients’ reports of their experience with illness are a key health outcome, but scores that measure these reports can be difficult to interpret. Skindex-29 measures skin-related quality of life reliably and validly. In this issue, Prinsen and colleagues compare patients’ responses to “anchor” questions with their Skindex scores to derive clinically meaningful scores for the subscales of Skindex-29. The cutoff scores identify patients whose skin diseases severely affect their quality of life.


Clinicians routinely interpret clinical measurements such as blood pressure or hematocrit: almost without thinking, we “know” what the values mean for a patient’s health. This knowledge is based on our experience caring for many patients and on what we have learned from teachers and reading.

For many measurements in clinical studies we lack this intuition, however. One of the original studies of finasteride for benign prostatic hyperplasia graphically illustrates this point (Lydic and Epstein, 1993; Gormley et al., 1992). Compared with placebo, the drug improved urine flow by an average of 3 ml/second—a rather bland finding, to be sure—until a subsequently published epidemiological study found that for men aged 40–74 typical urine flow rates decline approximately 0.2–0.3 ml/sec per year of life (Girman et al., 1993). We can now interpret the real-world effect of finasteride: on average, it restores a man’s urinary flow to what it was 10–15 years earlier.

Interpreting measurements such as urine flow is relatively straightforward compared with interpreting more abstract health outcomes such as scores of patients’ reports concerning their experience with illness. The actual measurement of psychometric constructs is a highly advanced and rigorous science, and substantial progress has been made in applying this science to clinical medicine (Both et al., 2007). Dermatology has lagged somewhat in rigorous studies to interpret the meaning of psychometric scores, however, and these scores remain unfamiliar to most researchers and clinicians. It is highly fitting that in what the Journal has designated the Year of the Patient (Bergstresser, 2010), this issue contains a good example of a study to facilitate interpretation of one measurement of patients’ experience of illness (Prinsen et al., 2010). In this Commentary, I describe briefly where we are with respect to measuring patients’ reports, summarize the major findings of that article, and project where we might go next to advance this important aspect of clinical research.

Patients’ reports of their experience with illness are a key health outcome. This observation is especially true for skin diseases, which do not typically affect survival, laboratory values, or easily measured clinical changes. In fact, patients’ reports are arguably an essential health outcome for dermatology because skin diseases (unlike most “internal” diseases) can change appearance, and they may have psychological and functional effects that cannot be assessed except through patients’ reports.

Most scientific work on the assessment of patients’ experience with cutaneous illness has focused on instruments that measure skin-related quality of life. Generic and disease-specific quality-of-life instruments have been developed...
Clinical Implications

- Tools to measure patients’ reports of their experiences with illness must measure accurately, and the results must be interpretable.
- Survey instrument scores are often not meaningful because they are unfamiliar and not empirically derived or clinically based.
- Information about interpretation of quality-of-life scores provides a basis for their more widespread use in clinical medicine and research.

for dermatology and found to have reliability (i.e., they give the same result when quality of life is the same), validity (i.e., they measure quality of life), and responsiveness (i.e., they change when quality of life changes) (Both et al., 2007). But fewer data exist on the interpretability of scores with these instruments. What does a given score mean? Does the score indicate severe effects of the disease or mild effects? What do changes in scores mean? Have the effects of the disease changed substantially or only by a small amount?

By simply examining the content of questions and patients’ responses, one can begin to interpret a score, especially for a single question. For example, a typical item in Skindex-29 is “I am embarrassed by my skin condition.” Response choices and corresponding scores are “never” (0), “rarely” (25), “sometimes” (50), “often” (75), or “all the time” (100). If a patient’s item score is 25, we understand that he or she is only rarely embarrassed by the skin problem. This “content-based” interpretation is less straightforward for scales that are derived from multiple items, however. Skindex is a multiscale index for which subscores are reported for Symptoms, Emotions, and Functioning. What can we do to put scale scores into context so that their meaning can be understood by clinicians?

A useful framework categorizes interpretation methods as either distribution-based or anchor-based (Lydick and Epstein, 1993). Distribution-based interpretations are based on the statistical distributions of scores in a given population. For example, I can begin to understand the magnitude of the effect my patients report by comparing their scores with those of a “normative” sample of unaffected persons (or of persons known to be severely affected). A recent paper using a distribution-based method reported that the distribution of responses to Skindex-29 could be clustered into statistically distinct categories based on the degree of reported quality-of-life effect (Nijsten et al., 2009). For the Symptoms subscale, for example, the categorization permitted cutoff values that corresponded to “very little” effect (≤3), “mild” effect (4–10), “moderate” effect (11–25), “severe” effect (26–49), and “extremely severe” effect (≥50).

Anchor-based interpretations, on the other hand, are made when scores are compared, or anchored, to other clinical results. A commonly used anchor is the response of patients to global rating questions that are themselves easily interpreted; in the current study, Prinsen et al. (2010) use this strategy to help interpret the meaning of Skindex-29 scores. The investigators administered Skindex-29 and a variety of anchor questions to a large sample of dermatology outpatients. The analyses compared Skindex subscale scores to patients’ responses to three major types of anchors: global questions about aspects of health-related quality of life, a question about the patient’s estimate of the clinical severity of his or her skin disease, and results on a standardized measure of psychiatric morbidity. For each of the anchors, the investigators predefined scores that indicate “severe” effects. They then determined the minimum Skindex scores (cutoff scores) that were most accurate in distinguishing patients who did or did not report severe effects. Skindex cutoff scores for severely impaired skin-related quality of life were ≥37 for Functioning, ≥39 for Emotions, and ≥52 for Symptoms.

Prinsen et al. (2010) used receiver-operating characteristic (ROC) methodology to determine the accuracy of the cutoff score. ROC curves are commonly used to display the ability of a diagnostic test to distinguish between people with or without the condition of interest by describing the performance of the test as the relation between the true-positive rate and the false-positive (1-specificity) rate (Deeks, 2001). Different cutoffs of scores have different sensitivities and specificities in relation to the criterion in question (e.g., global health-related quality of life). To determine cutoffs for Skindex scores, the authors selected the cutoff that maximized the sum of sensitivity and specificity (Fluss et al., 2005). This decision does not ipso facto have clinical meaning but requires a judgment about the relative benefits and liabilities of accurately detecting and not missing severe quality-of-life effects. With justification, the authors could also have chosen different levels of sensitivity and specificity (e.g., to maximize true positives at the expense of also increasing the numbers of false positives). Their strategy seems reasonable, however. Lowering the Skindex Symptom cut-off score for severe quality-of-life effects from 52 to, say, 45 would have detected more patients with severe Symptom quality-of-life effects as measured by the global item, but also would have labeled some patients as severely affected who in fact did not have severe effects as measured by the global item.

I look forward to seeing more results from this important and careful study, particularly the cutoff scores for mild and moderate degrees of effect, as determined by the anchors. Such results would permit us to interpret changes in scores if, for example, a group of patients changed from “moderate” quality-of-life effects to “mild” effects over time or after an intervention. They would also permit a more in-depth comparison to the Skindex-29 cutoffs derived from the distribution method described above (Nijsten et al., 2009). Although the results for the interpretation of “severe” quality-of-life effects are similar in the two studies for Symptoms and Functioning, the cut-off for severe Emotional effects in the current study (≥39) would be classified as indicating only moderate effects in the distributional study.
Using different anchors can inform interpretation even more. Prinsen et al. (2010) determined Skindex cutoff scores for patients’ judgments of severity of their disease and for their responses to a measure of psychiatric morbidity. In other work, changes in generic health-related quality of life have been correlated with the impact of stressfull life events, with being diagnosed with chronic diseases, with resource utilization, and with survival (Deyo and Patrick, 1995).

Clinically meaningful interpretations of quality-of-life scores are important to patients, clinicians, researchers, public health personnel, and policy makers. These individuals will be comfortable with these scores only when they become familiar (Deyo and Patrick, 1995), which will require their routine use in clinical research and possibly in practice (Chren, 2005). But routine use alone is not sufficient. Even if widely interpretable scores are obtained and reported in clinical trials, the results may not be used to modify the conclusions (Contopoulos-Ioannidis et al., 2009). Ultimately, to improve clinical decision making, an explicit commitment to including patients’ perspectives is necessary in clinical research.

CONFLICT OF INTEREST
The author states no conflict of interest.

ACKNOWLEDGMENT
This work was supported by National Institutes of Health grant K24 AR052667.

REFERENCES

Common Links among the Pathways Leading to UV-Induced Immunosuppression
Gary M. Halliday

In this issue, Sreevidya et al. demonstrate unexpected similarities in the downstream events of two molecular triggers of UV-induced immunosuppression. Both platelet-activating factor and cis-urocanic acid produced reactive oxygen species (ROS). Blocking both photoproducts reduced UV-induced genetic damage. Vitamin D, another immunosuppressive photoprodutct, does not share this property but instead enhances DNA repair. This study therefore links ROS and genetic damage with two molecular triggers of UV immunosuppression.

See related article on pg 1428