The Challenge of Understanding Articles about Health-Related Quality of Life

Ira Wilson
Tufts–New England Medical Center, Boston, Massachusetts

(See the article by Coplan et al. on pages 426–33).

Even without formal training in study design and biostatistics, clinicians can often use common sense, clinical judgment, and experience to help them decide which journal articles to pay attention to. Just as you do not have to be a professional baseball player to appreciate the subtleties of a well-played baseball game, most clinicians have an intuitive feel for what makes good clinical research.

On the other hand, for most clinicians (in this country), reading an article about health-related quality of life (HRQoL) is a bit like listening to cricket scores on British Broadcasting Corporation radio: you know that it involves a ball and a bat, and you strongly suspect that the people watching it are very civilized and drink tea. But that is about it. Even though interest in the measurement of HRQoL has grown over the past 10 or 15 years [1, 2], most readers of clinical journals have not yet developed an intuitive feel for the measures used to assess HRQoL.

Understanding HRQoL measurements is not as difficult as it might seem. Below, I highlight some issues to consider when reading articles that use HRQoL measurements and then use the article by Coplan et al. [3] in this issue to illustrate these issues. But before I do so, I would like to briefly discuss 2 things that can make understanding quality-of-life measurement difficult. The first relates to what HRQoL measurements hope to determine; the second relates to how the measurement is done.

What HRQoL measurements hope to capture are patients’ subjective perceptions and assessments of their health. These perceptions and assessments cannot be measured by blood testing, electroencephalography, MRI, or any other “objective” testing [4–6]. Physicians should be very comfortable with these subjective elements of health—they are what we strive to understand when we obtain a history. When a patient reports chest pain or fatigue, we try to assess (among other things) the severity of the symptom, what activities it interferes with, how the patient is adapting and coping, and how the symptom is affecting the patient’s overall sense of well being. Thus, HRQoL measurements attempt to capture a subset of the data that would be captured in a routine patient interview [7, 8].

These data exist on a continuum of increasing biological, social, and psychological complexity [9, 10]. The frequency and severity of chest pain is not conceptually complex. However, the general health perceptions of a person with chest pain might involve many dimensions of a person’s life, including other illnesses, threats to employment and social functioning, and worries about the future. Two people with chest pain of the same severity might, because of these other factors, rate their general health perceptions very differently. Clinicians know this instinctively and should use their clinical instincts when trying to understand articles that use measurements of HRQoL.

The “how” part of HRQoL measurement is necessarily a little more mysterious. HRQoL can only be assessed using survey research methods, in which few clinicians have formal training. There are measurement theories and mathematical models that underlie the approaches that survey researchers take to developing survey items and aggregating items into scores [11]. Unfortunately, when you read methods sections to learn about this measurement science (called “psychometrics”), you see terms like “Cronbach’s alpha,” “construct validity,” and “principal components analysis.” These are not inviting phrases to the inquiring clinical mind. The important message is that HRQoL can be measured validly, and interested readers are referred to relevant texts [12].

What are some of the issues that clinical readers should consider when reading articles about HRQoL?
1. Is there some theory of health underlying what the authors measure?

There is sound theory and many years of careful work that underlie the approach taken by Coplan et al. [3], although this earlier work is not referenced or described [13, 14].

2. Are the measurements reliable, valid, and responsive?

Reliability refers to the extent to which a measurement is free of random error [15]. Validity is the extent to which a measurement measures what its developers intend it to measure. Responsiveness is the extent to which the instrument can detect clinically important changes in health. Readers interested in these basic measurement concepts are referred to other sources [5, 11, 15]. Coplan et al. [3] appropriately refer readers to articles that address the reliability and validity (references 6 and 7 in their article) of the items and scales they use. The authors do not discuss responsiveness.

3. Do the items and scales make sense clinically?

I argued above that measurements of HRQoL assess many of the same things that clinicians regularly ask patients about, and I encourage clinicians to read the items that make up HRQoL scales and make a clinical assessment of their meaning. To find the actual survey that Coplan et al. [3] used, I had to do an Internet search (https://www.fstrf.org/qol/aactg/adult_ql.html). I would encourage authors to include such links to actual surveys in their articles. Only by looking at the actual items can you begin to understand, for example, why results might be different for similarly sounding outcomes like “general health” and “current health perceptions” (see figure 2A in [3]).

4. Is the time frame clinically sensible?

Twenty-four weeks is probably an appropriate time frame for the assessment of some outcomes of antiretroviral therapy. However, it is not clear what time frame is optimal if the intent is to capture and understand the metabolic and bodily changes that can accompany antiretroviral therapy. Some of these complications may take >24 weeks to develop [16].

5. Is some attempt made to help readers understand “effect size”?

Coplan et al. [3] highlight the finding that persons receiving triple-drug combination therapy including indinavir had improvements in general health of 2.9 points, compared with a decrease of 0.2 points for persons receiving dual-nucleoside therapy \( (P = .018) \). Because this was a relatively large trial \( (n = 1156) \), the findings were statistically significant, but how are readers to understand whether this difference is clinically important? Others have articulated this problem as understanding the “minimally clinically important difference” in a given scale [17–19].

6. Are links made between clinical measurements and measurements of HRQoL?

The causal model implicit in this paper is that effective antiretroviral therapy causes reductions in virus loads, which in turn causes increases in CD4 cell counts, which in turn improves HRQoL. Clinical readers are used to thinking about changes in virus loads and changes in CD4 cell counts. How are changes in general health related to these more familiar virological and immunological measures? Readers would understand HRQoL measurements better if the relationship to more familiar clinical variables were demonstrated [9, 20].

Coplan et al. [3] have conducted a careful, thoughtful analysis and are appropriately cautious in the interpretation of their results. Their findings help clinicians understand that triple-drug combination therapy that includes indinavir certainly does not reduce HRQoL and may improve it slightly, compared with dual-nucleoside therapy. No one article can be expected to address all of the issues that I have suggested that clinical readers think about when they read an article about HRQoL. However, the more of these issues that are addressed, the better clinical readers will be able to understand and apply a study’s findings.

The final observation I would make is that, to my knowledge, this is the first published article from the AIDS Clinical Trials Group that uses these specific HRQoL measurements, despite the fact that they have been in use in clinical trials since 1997. Although I applaud the publication of this paper, it is unfortunate that these findings are coming to light nearly 7 years after publication of the study’s primary results. HAART, although effective, has many toxicities, and timely analysis and publication of HRQoL data from clinical trials that have collected these data is one of the best ways to understand these toxicities.

Acknowledgments

Conflict of interest. I.W.: No conflict.

References


