Published in final edited form as:

J Soc Integr Oncol. 2006 ; 4(2): 100–103. doi:10.2310/7200.2006.007.

How to measure quality of life in integrative oncology research

Andrew J. Vickers^{1,2}

¹Department of Medicine, Memorial Sloan-Kettering Cancer Center

²Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center

Abstract

Quality of life is an important focus of integrative oncology research. Choice of an appropriate way to measure quality of life is therefore a key concern. A common error of researchers has been to focus on whether a quality of life measure has been "validated". This may be a necessary condition for choice of endpoint, but it is certainly not sufficient. It is easily shown that interventions that undoubtedly improve quality of life, such as Epogen for chemotherapy anemia, will not have an important impact on items in some validated quality of life scales (e.g. "I trust my doctor") and therefore will not improve quality of life scores. Investigators are advised to move beyond "validation" as a criterion for choosing quality of life endpoints and focus instead on the specific content of different potential measures. A particularly important concern is whether a prior trial showing a statistically significant difference on a specific quality of life scale is published, as this would demonstrate that scores on the scale will improve if patients respond to treatment. In many cases, it is preferable to use specific symptoms, such as pain or fatigue, as opposed to more generic quality of life scales.

Introduction

Quality of life is an important focus of integrative oncology. Many integrative oncology interventions, such as massage, acupuncture and music therapy, have no effect on cancer cells and are used because they help patients "feel better". Research on such interventions therefore involves defining and measuring how and whether patients "feel better" after a particular treatment. This typically has fallen under the rubric of "quality of life research". A key aspect of such research is the appropriate selection of an instrument, such as a questionnaire, to measure of quality of life. In this paper, I will outline some of the pitfalls of quality of life measurement and suggest guidelines for trial design, particularly with respect to randomized trials. I aim to be informal and didactic so I do not give statistical examples or references: these are available on request.

Measuring quality of life: the naïve syllogism

I review a good number of integrative oncology trials, either as proposals or published trials. In the vast majority of cases, the authors justify the choice of study endpoints, such as a quality of life questionnaire, by claiming that they have been "validated" by prior studies specially designed to study their properties. In other words:

- Questionnaire X is a validated quality of life measure
- My intervention Y improves quality of life

And therefore:

- Intervention Y will improve scores on questionnaire X

To explain why I do not find this argument to be valid, consider the “Best Places to Live” lists published by several magazines. These rank towns and cities by quality of life indicators such as income, housing costs, housing size and commute time. Now I have little doubt that my own personal quality of life would be improved if I earned more, spent less on my mortgage, had a bigger house and spent less time getting to work. The point is, however, that a course of massage treatments will make no difference to any of these. Yet no-one would argue on these grounds that massage does not improve quality of life.

To illustrate this point further, I chose four interventions that I believe have important effects on quality of life: acupuncture for chronic pain, Epogen for anemia related to chemotherapy, psychotherapy for depression and exercise during chemotherapy. I then selected four individual questions from several quality of life questionnaires: “I feel that my family supports my treatment decisions”; “I trust my doctor”; “I can walk up several flights of stairs”; “I feel sick or nauseous”. It is clear that these are important issues that affect quality of life. For example, my quality of life would be poorer if I were nauseated, unable to walk up stairs or if I had a life-threatening disease and was unable to trust my doctor.

In table 1, I pair each intervention with one of the quality of life items. If we conducted a randomized trial, and compared the responses of the quality of life item between intervention and control, I doubt we would find any substantial differences between groups. There might be some minor effect: for example, it may be that some depressed patients feel so fatigued that they are unable to climb stairs; similarly, better well-being during chemotherapy could possibly improve the trust between patient and doctor. However, clinical research is a crude tool, and it is not good at identifying subtle differences between groups. When designing a randomized trial, investigators should choose an endpoint with the greatest possible difference between groups. In this respect, acne research gives us some important lessons.

How acne research can help us design trials in integrative oncology

Acne research provides an excellent case study of endpoint assessment in quality of life trials. First, acne does not shorten life and acne treatment does not extend it, hence the purpose of any trial of an acne therapy, like many of those in integrative oncology, is to determine whether treatment improves quality of life. Second, the endpoints used in acne research are relatively standard. Third, in several trials, outcome has been assessed using multiple endpoints so that these can be compared. Some of these endpoints are described in table 2.

To interpret the results of acne research, we have to bear in mind what a typical acne treatment is meant to do. Acne is a quality of life issue, but acne treatments are not designed to improve quality of life, they are designed to treat and prevent skin infection. A highly effective acne treatment will reduce the number of infected comedones (spots), and hopefully this will improve the patient’s facial appearance. This is likely to ameliorate anxiety and depression both directly, and through the indirect result of an improved social and romantic life.

But there is nothing inevitable about such quality of life improvements: a patient may have fewer spots but still dislike his facial appearance, or still suffer from other skin conditions, or other ailments, or just be generally unhappy with his life. In table 2, the d refers to the effect size – the difference between treatment and control groups divided by the standard deviation – for various endpoints after a successful acne treatment. The higher the d , the more likely any particular trial is to find a statistically significant difference between groups. As one moves from the symptom, which is the specific problem the treatment is designed to address, to

increasingly general aspects of quality of life, d decreases. Thus, the overall lesson from acne research is: work out what the intervention is meant to do and measure that; more general effects on quality of life would be nice, but are likely to be smaller in degree and therefore more difficult to detect in randomized trials.

I saw a similar effect in one of my own studies. In a randomized trial of acupuncture for chronic headache disorders, we saw a highly significant difference between acupuncture and control with respect to headache pain ($p = 0.0002$). Now we would expect that a patient who experiences less headache pain after acupuncture treatment will also experience improvements on quality of life dimensions such as physical functioning, social functioning, general health and energy. We measured several of these domains using the SF-36 health status scale, and although we found that patients in the acupuncture group did generally do better than those in the control group, differences between groups were not always statistically significant. Moreover, where statistically significant differences were found, p values were 0.02 or 0.036, not the highly significant result we saw for the headache pain itself. In other words, acupuncture was better at doing what it was supposed to do (reduce headache pain) than at bringing about general improvements in quality of life.

Therefore, my first piece of advice for choosing a quality of life endpoint for a clinical trial is as follows. First, identify a number of different questionnaires that are plausible candidates for the endpoint in your trial. A good place to start would be to identify some trials similar to the one you plan to conduct and then examine the endpoints used. Second, obtain actual hard copy of the questionnaires. Third, read every question on each questionnaire and consider whether your study intervention will have a *large* effect on how a patient might answer it. Fourth, choose the questionnaire that has the questions most relevant to your intervention and patient population.

The Integrative Medicine Service at MSKCC conducts a large number of trials. We do not commonly use general quality of life measures at all. Instead, we measure specific symptoms. For example, for an ongoing trial of acupuncture for pain after surgery for lung cancer, we had asked our acupuncturists to design a treatment specifically to treat pain, and it is reasonable to assume that differences in pain will be the largest effect that we will see. The endpoint we are using is therefore the Brief Pain Inventory, which asks questions about the level of pain and how much it interferes with everyday life. A patient with an average pain score of, say, 2 out of 10 is likely to have a better quality of life than one who scores 7 out of 10; however, I am not sure we really need to prove this by asking patients to complete a generic quality of life questionnaire.

What is meant by "validation" of a quality of life scale?

As mentioned above, the typical quality of life paper or protocol will describe, often in some detail, the "validation" of the scales used in the trial. This is somewhat like choosing a restaurant for a birthday treat on the basis of its food safety certificate: nice to know, but hardly the key issue. One particular reason why we do not use food safety as a criterion for restaurant selection is that it does not discriminate very well, that is, there are numerous restaurants and the vast majority of them are certified safe.

In some senses, validation of quality of life trials is somewhat worse than food safety certification. State departments of health have very specific criteria for determining whether a restaurant meets standards and give each a simple pass or fail rating. Validation studies, on the other hand, typically report correlations between a new scale and either a more established scale or a clinical category, such as disease stage. They may also report measures such as how much individual scores change from day-to-day.

The problem is this: how high a correlation is high enough, and what level of variation is low enough, to make a scale valid? I don't think anyone really knows and as a result, investigators in a validation study almost always report that correlation and variation were acceptable and, accordingly, conclude that the scale has been validated. As a short demonstration, I searched Medline for *validation*, *"quality of life"* and *scale* on February 2, 2006. I then documented the key sentence from the conclusion of the first 15 papers I retrieved, omitting two papers that were not relevant. The results are shown in table 3: every single study concluded that the instrument under study was "valid".

I propose an alternative to citing validation studies to justify the choice of quality of life scales in clinical trials. First, choose a quality of life scale that you think would be appropriate for your study. Second, look to see if previous clinical trials have been published that used the scale as an endpoint and reported a statistically significant result. And that's pretty much it. All you really need to know is that, if the patient gets better, the quality of life scale reflects that improvement.

Conclusions

Measurement of quality of life is a key issue for integrative medicine research. As such it is disappointing that researchers have paid rather superficial regard to the choice of quality of life endpoints. In my view, researchers should not merely to refer to a "validation" study as sufficient justification for a choice of measure. Instead, investigators should focus on the specific content of different possible measures. A particularly important criterion is whether a prior published trial shows a statistically significant difference on a specific quality of life scale, as this would demonstrate that scores on the scale will improve if patients respond to treatment.

Acknowledgements

This research was funded, in part, by a P50-CA92629 SPORE from the National Cancer Institute.

Table 1

Example interventions matched with example items from quality of life scales. Although the interventions improve quality of life, and the items measure quality of life, the interventions will not have an important impact on how patients respond to the items.

Acupuncture for chronic pain

“I feel sick or nauseous”

Epogen for chemotherapy anemia

“I trust my doctor”

Psychotherapy for depression

“I can walk up several flights of stairs”

Exercise during chemotherapy

“I feel that my family supports my treatment decisions”

Table 2

Endpoints in acne research. Effect size, d , is lower for more general measures than for endpoints more closely related to acne treatment.

Number of infected comedones
Appearance of face
Dermatology life quality index
Health-related quality of life
Overall quality of life

d

Table 3

Conclusions from 15 recent studies examining whether particular quality of life scales were "valid".

1.	The excellent psychometric properties of this simple gastro-oesophageal reflux disease diary make it an attractive measure for future trials.
2.	The Global Overall Symptom scale is a simple, valid outcome measure for dyspepsia treatment trials.
3.	The results provide evidence of the good psychometric properties of the new HeRQoLED questionnaire,
4.	These results support that the EORTC QLQ-C30 (version 3.0) has proven to be a reliable and valid measure of the quality of life in Thai patients with various types of cancer
5.	The AAQoL appears to be a valid measure of quality of life for adults with ADHD and can be considered for incorporation into future studies
6.	The CAMPHOR is a valuable new instrument for assessing patient-reported outcome in PH clinical trials and routine practice
7.	The PROCOG demonstrated good to excellent psychometric properties among a sample of older adults
8.	The short-forms represent a valid substitute for the full scales, as they provide over 85% of full-scale information and demonstrate excellent reliability statistics.
9.	The modified MOS scale was reliable and valid for evaluation of QOL in HIV infected persons in India.
10.	This data suggests that the EORTC QLO-C30 and the QLQ-H&N35 are reliable and valid questionnaires when applied to a sample of head and neck cancer patients in India.
11.	Although the fit to the GPCM supports the construct validity of the Qualeffo-31, validation in a new study should be performed before using it in practice.
12.	The GISSK is a reliable and valid measure of GI symptoms and their severity in JRA
13.	A 45-item refined version of CADIS (score 0-180) [gave] evidence of content and construct validity and [is] suggested [for] use in clinical research
14.	Preliminary evidence is provided that the HRQOL scale could serve as part of program evaluation, intervention, and surveillance efforts among college students.
15.	The RLSQoL is a valid, reliable and responsive measure of quality of life for patients with RLS, in a clinical trial setting where group comparisons are anticipated.