

# ASSESSING THE CLINICAL SIGNIFICANCE OF QUALITY OF LIFE (QOL) MEASURES

**Jeff A. Sloan, Ph.D.**  
Quality of Life Research, Mayo Clinic  
Rochester, Minnesota



**ABSTRACT:** *Quality of life (QOL) impacts patients' well being and survival, yet many issues in QOL measurement exist. This article describes the state of the science in QOL assessment. The necessary complexity of QOL assessment is explained. Recommendations for choosing QOL tools are provided. Clinically significant change in QOL is defined. Approaches to incorporate QOL assessment into clinical trials are highlighted.*

QOL research can seem daunting, yet at it's simplest level, it is really about asking the patient the simple question "How are you doing?" Most scientific applications, however, require something a bit more sophisticated. While QOL can be expressed simply on a scale of 0 to 100 (0 is the worst and 100 the best QOL), consistent analysis of this scale (i.e., is it reliable and reproducible? Will it change?) is a key issue.

QOL research has a long history. In 400 BC, Hippocrates said that physicians should look after the quality and quantity of a patient's life. In 1935, Dr. William J. Mayo, one of the founders of the Mayo Clinic, said, "It is worthwhile to secure the happiness of the patient as well as to prolong his life."

Evidence in the literature suggests that items related to QOL, such as fatigue, social support, symptom distress, and group counseling are related to patient well being and have been proven to be correlated with patient survival. We asked advanced cancer patients in a North Central Cancer Treatment Group clinical trial about their QOL on a scale of 0 to 100. Patients reporting good QOL lived, on average, three months longer than those with lower QOL.

Correctly measuring and interpreting QOL is involved. We should not expect QOL to be any simpler or any more complex than any other clinical endpoint. Measuring seemingly straightforward endpoints, such as response to treatment can even be complex. For example, according to the RECIST criteria for tumor response assessment mandated by the National Cancer Institute, as many as 10 indicator lesions may be included in the determination of response, with a table to further outline response categories. Estimates for complete and partial response are often off by as much as 20%. Why would we expect measuring QOL to be simple when measuring tumor response is complex?

Realistically, one can think of QOL as just another surrogate endpoint biomarker (SEB). Even though mechanisms and measures are still needed, for many SEB's, such as genetics endpoints, there is a greater general acceptance for SEB's than QOL endpoints. For example, more than 10,000 articles have been written on the P53 marker, which is related to an increased risk of breast cancer; however, there is as of yet no clinical indication for this marker. QOL has an advantage over many SEB's in that there are obvious interventions available for QOL endpoints.

Articles on fatigue recently published in Oncology demonstrated that fatigue is related to patient survival. Mechanisms of action and treatments to reduce fatigue and, hopefully, improve patient survival have been identified.

There is an assumption that QOL endpoints are "soft" and therefore require a higher standard. Part of the issue here is familiarity with the endpoint. The measurement of blood pressure via a cuff faced similar criticism 100 years ago. Today, virtually all clinical trials regularly take patients' blood pressure, yet 100 years ago, measuring blood pressure by a machine was a new idea. Trials were started to determine whether this new machine could actually measure blood pressure. At that time, massage therapy was believed to alter blood pressure, so researchers designed experiments in which they used the new machine to measure blood pressure, gave the patient a massage, and then used the machine again. If the machine changed its number, this would be accepted as evidence that the patient had a clinically significant change in blood pressure.

One hundred years later, trials are underway to determine if massage therapy can help alleviate anxiety, pain, and other psychosocial

correlates of cancer. These trials use the blood pressure cuff to see whether massage therapy has an impact, because the blood pressure cuff is now the standard. The irony of this scientific circularity illustrates why QOL is still viewed with some skepticism.

QOL measures are at the point blood pressure measurement was 100 years ago. Blood pressure measurement has become accepted because we are familiar with it. Through experience and experimentation, we know the clinical indications. We know what 140/70 means. We do not have this knowledge yet in QOL. Many measures are relatively young. For example, the first QOL measure specific for cancer patients, the Functional Living Index of Cancer, was first published in 1986. With time, QOL can become as routine a clinical indicator as blood pressure is today.

### Choosing the Best Tool

The medical profession is waiting on and looking for a QOL assessment that is easy to understand, use, and be completed in less than five minutes. However, it is hard to get a really good description of a patient's QOL in a single question. The best QOL tool is the one that answers the question. Regulatory agencies, cooperative groups, and investigators are struggling to determine the best tool for assessing QOL. Recommendations are now coming out in the literature in terms of clinical significance, but they are slow in coming. It is like prescribing prunes for constipation: one might not be enough but 50 is definitely too many. An 18-page QOL questionnaire is probably overkill.

One long-standing debate in the literature is whether one question, a single item, is enough. If you do not need a great deal of detail in assessing QOL, a single item might fit - it depends on the research, one size does not fit all. It is not an

adversarial situation. There may be some aspects of QOL, for example, fatigue and pain, where "how are you doing?" is adequate. In other situations, such as measuring psychosocial well being, you may want a more complete psychological profile. Identification of the study and researcher's needs allows us to delineate when single or multiple items are appropriate.

Often, the single item will be sufficient in large population studies where researchers want a basic idea of QOL. However, QOL is multi-dimensional and multi-faceted, so you probably need multiple items. More detail and more depth also help you generate more science. As the complexity of the research increases, you need more QOL items, but there is a saturation point where researchers and patients want "bite size morsels" about QOL.

In general, we have found that "less is more" in QOL assessment. Single items have greater variability, sensitivity, and ease of application. They can also be used as triggers for further exploration. For example, we have found that, on a QOL score of 0 to 10, people who give a score of 7 or higher are likely not in need of further assessment, but people who give a score of 6 or lower have something they want to talk about. A score of six or lower is an indication for clinicians to further question patients.

We recently published guidelines on assessing QOL by starting with a simple question, "What is your overall QOL?" If patients say 8 to 10 we move on. If they say six or seven, they might need help, and we ask five general questions (e.g., what is your physical QOL like? What is your intellectual QOL like?). If we see an issue, we deal with it. If patients give a score of zero to five we need to do more work, possibly using a multi-dimensional questionnaire that examines in detail the various aspects of QOL.

The basic idea is to keep it simple (Table 1). A QOL assessment with fewer than 25 items will probably provide complete data with few issues related to completion rate. With 25 to 50 items, people start getting tired. With more than 50 items, you will lose people. With more than 100 items, people will get angry. The number of items incorporated should be carefully considered. CRAs must let investigators know what patients do not like about QOL assessments (e.g., patients think asking 10 questions about their sexuality has nothing to do with their cancer).

It is vital to get input from the people who will handle the data before the study begins. QOL assessment should be treated like other data. Include some data checks, but no more than for other endpoints. Include the reason for

**TABLE 1**  
**Keeping QOL Measures Simple**

Length of QOL measures:  
Less than 25 items: produces complete data and few issues  
25-50 items: people start to get tired and missing data begins to appear  
50-100 items: you will lose people and entire forms are left blank  
More than 100 items: people get angry  
Use uniform response categories (e.g., Yes/No)  
Design measures that look professional  
Get input from all people who will handle the data BEFORE the study opens

missing data. One of the classic situations is when you ask, “do you have pain” and a follow-up question, “when you have pain, what is the intensity of the pain?” If the patient does not have pain, that question will be blank. All we need to know is that the patient left it blank because it does not apply; that the data are not missing. If patients became ill halfway through the study, there should be a place to write this information on the questionnaire.

Often, the reason for missing data is informative. For example, in one study we had a treatment that caused people to gag. When we asked people what their swallowing was like, we got a lot of missing data other than “couldn’t swallow.” When we compared the placebo and control groups, the placebo group had few missing data and the control group had many missing data on this issue.

It is useful to identify a CRA to be responsible for coordinating QOL data management. This improves the consistency and data quality tremendously. Additionally, QOL assessments should look like professional surveys, with instructions at the front end, and the picture, name, and phone number of the investigator for any problems that may arise. Patients like the opportunity to call, although very few actually do.

**General Classifications for Methods Assessing Clinical Significance**

While analyzing QOL data, it is important to recognize that a statistical change does not always imply a clinically significant change. In determining clinically significant changes, some say there are no set rules and only a group of concerned, competent, and experienced people can determine whether a clinically significant change has occurred. A simpler view that is emerging in the literature is to say, “If it looks like a duck, sounds like a duck, and walks like a duck, then the odds are remote

<b>TABLE 2</b>
<b>Methods for Assessing Clinical Significance in QOL Assessment</b>
Tool-specifications (norms, experience)
Investigator-defined (Effect Size, Standard Error of Measurement, Empirical Rule of Effect Size, R-Squared Methods)
A posteriori patient-defined (Minimally Clinically Important Difference, Minimally Important Difference)
A posteriori statistically-defined (emergent p-value)
Anchored to clinical outcome (e.g., ability to walk); this is probably redundant

of it being a worm or an elephant in a clever disguise.”

The literature includes many methods of assessing clinical significance (Table 2). There are tool specifications (norms and experience) and investigator-defined methods of assessing clinical significance. Another arena of clinical significance has been called the minimal clinically important difference (MCID) or minimally important difference (MID), which asks patients how their QOL has changed since their last physician visit. To determine a change, you look at the distribution of patients who said their QOL has changed and determine how much QOL needs to change to be perceived. However, the emergent p-value is definitely not a good way of defining clinical significance. Another method of assessing clinical significance is to anchor it to a clinical outcome, for example, if the patient can now walk, his/her QOL has improved. However, if the clinical outcome is that which is of interest, and not the QOL score, gathering or looking at a score is unnecessary.

Clinical significance boils down to whether the amount of change in patients’ QOL score is so huge that researchers cannot ignore it, so small they would never do anything about it, or somewhere in the middle. All of the methods for assessing clinical significance outlined above converge to the same conclusion that a movement equivalent to 1/2 standard deviation is a minimally required shift for

<b>TABLE 3</b>
<b>A Proposed Unified Theory Approach for Assessing Clinical Significance in QOL Assessment</b>
All methods can be equated
Define a movement equivalent to ½ standard deviation as a minimally required shift for clinical significance on any domain or individual item
Define protocol-specific exceptions a priori with support evidence (e.g., 1/4 standard deviation is important here because...)

clinical significance on any domain or individual item (Table 3). We now have a guideline for assessing the clinical significance of any QOL tool: on a 0 to 100 point scale, a change of 10 points is clinically significant.

In summary, while QOL is still relatively new as an endpoint in clinical trials, experience with it is growing. Time and familiarity with the basic methods will eventually make QOL as routine as other clinical endpoints. We are presently touching the most difficult issue of determining a clinically significant effect in QOL. A picture is already emerging, however, that while it may first seem complex, it can be made simple. The CRA has a critical role in ensuring that high quality data are obtained, for without it, the QOL of the investigator is put in jeopardy.