



Published in final edited form as:

J Soc Integr Oncol. 2006 ; 4(3): 135–138. doi:10.2310/7200.2006.017.

Multiple assessment in quality of life trials: how many questionnaires? How often should they be given?

Andrew J. Vickers^{1,2}

1 Department of Medicine, Memorial Sloan-Kettering Cancer Center

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

Abstract

Researchers conducting randomized trials of integrative interventions often ask patients to complete multiple different quality of life questionnaires or symptom severity scales repeatedly over the course of a trial. Although trialists rarely if ever give a strong justification for either the number of questionnaires they give or how often they give them, these are design decisions that can be taken systematically. Giving large numbers of questionnaires can improve the precision of trial results, and provide interesting secondary data, such as on the time course of symptoms. However, doing so can also lead to excessive patient drop-out, an undue data management burden and difficulties with interpretation of results. As a general guideline, each aspect of quality of life should be measured by a single questionnaire, and researchers should avoid giving more than three different questionnaires to patients. Decisions about the appropriate number of assessments to use can be based on statistical properties derived from simple formulae.

Introduction

Researchers conducting randomized trials of integrative interventions often ask patients to complete multiple different quality of life questionnaires or symptom severity scales repeatedly over the course of a trial. As a typical example, patients might complete the Profile of Mood States, the Hospital Anxiety and Depression Scale, the Beck Depression Inventory, a pain scale, a fatigue scale and the Functional Assessment of Cancer Therapy at baseline, every week during an eight-week treatment, and then monthly for a six month follow-up. In my experience, trialists rarely if ever give a strong justification for either the number of questionnaires they give or how often they give them. A trial protocol will commonly include a description of the psychometric properties of each questionnaire but not explain why more than one is required. Similarly, a protocol will generally merely state when questionnaires are to be administered, but not indicate how the number and timing of assessments was determined.

My principal point in this didactic paper is that both the number of different questionnaires and the frequency of administration are design decisions that need not be taken arbitrarily: trialists should avoid basing such decisions simply on “what seems reasonable” and use, instead, some simple guidelines and analyses. I will then outline one approach to determining the appropriate assessment strategy. However, the exact details of my proposed method are less important than the recommendation that some sort of systematic approach should be used when designing randomized trials incorporating multiple measurement.

Multiple measurement 1: measuring more than one aspect of quality of life

The simplest way to measure quality of life in a randomized trial would simply be to ask patients to give an overall rating on a 0 – 10 scale. There are several reasons why we avoid this approach. The first, which was discussed at some length in the prior article in this series(1), is that interventions typically affect some aspects of quality of life but not others. An analgesic drug will markedly reduce pain, but effects on overall quality of life are diluted by changes in other quality of life domains unrelated to pain. Hence we often choose to study specific symptoms or domains - pain, depression, physical functioning and so on - than use a global measure. We might also avoid a single quality of life rating because we want to explore the effects of an intervention on different aspects of quality of life, such as the relative degree to which physical and psychologic domains improve after treatment. As a concrete example, I am currently working on a trial comparing massage to psychotherapy in patients with advanced cancer. The investigators hypothesize that while massage may reduce anxiety scores, it will have a smaller effect than psychotherapy on existential concerns, as measured by a spiritual well-being scale.

The key point here is that although we might want to measure different quality of life domains, there is no need to measure a single domain in different ways. Look again at the example used in the introduction, a trial in which patients completed the Profile of Mood States (which gives separate scores for anxiety, depression, hostility, fatigue, vigor and confusion as well as an overall mood disturbance score), the Hospital Anxiety and Depression Scale, the Beck Depression Inventory, a pain scale, a fatigue scale and the Functional Assessment of Cancer Therapy (which rates for physical, functional, social and emotional domains). As a result, the researchers measured depression in three different ways; anxiety, fatigue and overall mood were assessed using two different scales.

There are several problems with this type of redundancy. The first is that the more you ask patients to do, the less likely they are to do it. Patients asked to complete a large number of questionnaires at the end of a study may be unwilling to do so, leading to missing data. This reduces the power of a trial and can introduce bias if, for example, patients who do poorly tend not to complete questionnaires. Against this point of view, I have heard it argued that “it only takes 2 – 3 minutes to fill in a questionnaire, so giving patients 6 questionnaires rather than 3 only entails an extra 5 - 10 minutes; I don't think 5 minutes of someone's time is going to make that much difference”. It is my experience, however, that sick patients given a stack of questionnaires do not make decisions based on formal estimates of the time commitment required. Rather, there is an immediate emotional reaction to seeing page after page of questions that can turn a patient against providing any data at all.

Increasing the number of questionnaires also increases the burden of data management, data entry and quality assurance. This might sound a trivial point, but in the initial example given in the introduction, where six questionnaires are given on 15 occasions, there are approximately 2200 data points per patient, or nearly a quarter of million data points for a 100 patient trial.

The most important problem of multiple questionnaires concerns interpretation of results. It is a useful exercise when designing a trial to identify possible different results and then consider how each should be interpreted. The problem comes if a single quality of life domain is measured in two different ways with discordant results. For example, what conclusions should be drawn if an intervention improved depression as measured by the Hospital Anxiety and Depression Scale but not Beck Depression Inventory scores?

Accordingly, I propose a simple strategy to decide how many different questionnaires to give.

1. Choose a list of possible questionnaires following the guidelines outlined in the previous article in this series(1). In brief, this involves choosing questionnaires that will show large changes following treatment.
2. Identify the quality of life domains or symptoms that are likely to be strongly affected by treatment.
3. Choose a combination of questionnaires such that each domain or symptom identified in step 2 is addressed by only a *single* questionnaire.
4. Consider reducing the number of questionnaires if you have more than three: it is better to get no information on some endpoint of questionable importance and excellent data on the primary endpoint, than to get data of moderate value for both.

So for our example study I would suggest using the Profile of Mood States and the pain scale. I might be talked into also including the Functional Assessment of Cancer Therapy, but the Beck, fatigue scales and Hospital Anxiety and Depression scale are redundant.

Multiple measurement 2: repeat measurement

In the simplest scenario, we measure quality of life or symptoms at the beginning and end of treatment. One reason why we might instead choose to assess outcome at several different times is to assess the course of treatment effects. For example, in a trial of a relaxation therapy class versus standard care for chronic cancer-related anxiety, it would be interesting to know whether patients experience an immediate effect of treatment, or whether anxiety decreases only slowly over time, with increasing practice of the relaxation technique. It might also be nice to know whether the effects of treatment persist after the class, and if so, for how long.

I would question this rationale for repeat measurement. Again, I rely on the principle that the more you ask patients to do the less likely they are to do it, and on the principle that it is silly to compromise your data on the primary endpoint in order to get data to address a quite secondary question.

The more important rationale for repeat measurement concerns error: ask someone how they are doing just once and they might just be having a bad day; asking them how they are doing several times and you are more likely to get a fair picture of how they truly feel. This raises the question of how many times is enough. The key issue here is variation. A patient's weight does not vary much from day to day and so it is reasonable for a trial of an obesity treatment to weigh only before and after treatment. Mood is more variable and so we might want to measure mood several times. The most extreme variation is seen with an episodic condition such as headache: a patient might be pain-free one day and completely incapacitated the next. Accordingly, headache studies often ask patients to complete pain diaries three or four times a day for many weeks.

The formulae describing the effects of multiple measurement on trial characteristics were originally derived by Frison and Pocock in the early 1990's(2). This paper was rather technical, and general in application, so a few years ago I wrote a simplified introduction to the topic that focused specifically on trial design(3). In brief, I argued that there is a tension between minimizing measurement error and maximizing patient compliance. If we ask patients to complete questionnaires every day for many months we will indeed get a very accurate picture of their quality of life, but we are also likely to see patients drop-out; on the other hand, if we give only a single questionnaire, we can rely on most patients to complete it but, as pointed out above, a patient might be having an unusually good or bad day and our results will be prone to error. What I recommended was using some simple formulae to calculate the variance of a study endpoint when it is measured once, twice, three times and so on. The variance is

proportional to the sample size, that is, a 10% decrease in variance means a 10% decrease in the number of patients required. A researcher could then decide whether the decrease in sample size resulting from taking an additional repeat assessment would offset any increase in drop-out associated with the greater reporting burden on patients.

Table 1 shows how sample size requirements change when the number of baseline and post-treatment assessment varies (note that these results were based on a particular correlation between measures that, although common, may not reflect the data of any specific planned trial: investigators are referred to the original paper(3)). It is clear that repeat measures are subject to the law of diminishing returns: measuring outcome twice rather than once reduces sample size requirements by 20%, but measuring 8 times rather than 7 reduces sample size by only about 1%. Also, as might be expected, it is more efficient to repeat measurement at follow-up than at baseline. Finally, note that table 1 does not, by itself suggest any particular number of baseline and follow-up measures. If an endpoint is not burdensome, say, a simple 0 – 10 scale, it might well be worth measuring 8 times rather than 7: this would not be the case for a lengthy quality of life questionnaire.

Conclusions

Randomized trials with quality of life endpoints often involve the administration of several different questionnaires at numerous time points. The number of questionnaires and repeat assessments are design decisions that can be taken systematically. Giving large numbers of questionnaires can improve the precision of trial results and provide interesting secondary data, such as on the time course of symptoms. However, it can also lead to excessive patient drop-out, an undue data management burden and difficulties with interpretation of results.

As a general guideline, each aspect of quality of life should be measured by only one questionnaire, and researchers should avoid giving more than three different questionnaires to patients. Decisions about the appropriate number of assessments to use can be based on the statistical properties derived from simple formulae.

Acknowledgments

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

References

1. Vickers AJ. How to measure quality of life in integrative oncology research. *Journal of the Society for Integrative Oncology*. 2006in press
2. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med* 1992;11(13):1685–704. [PubMed: 1485053]
3. Vickers AJ. How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Med Res Methodol* 2003;3:22. [PubMed: 14580266]

Table 1**Reduction in sample size associated with increasing the number of repeat assessments in a randomized trial**

The results are based on the common situation where the correlation between baseline measures is 0.7, between follow-up measures is similarly 0.7 and between baseline and follow-up measures is 0.5. It is assumed that groups will be compared by analysis of covariance, using the mean of follow-up measures as the outcome.

Change from		Change to		Sample size reduction
No. of baseline measures	No. of follow-up measures	No. of baseline measures	No. of follow-up measures	
1	1	1	2	20%
		1	3	27%
		1	4	30%
		1	7	34%
		1	8	35%
		4	1	10%
		4	4	40%
4	4	4	7	7%
		7	4	3%
		7	7	10%
7	7	7	14	5%
		14	7	2%
		14	14	8%
14	14	28	28	4%