THE MEASUREMENT OF QUALITY OF LIFE IN CLINICAL TRIALS: CONSIDERATIONS OF GOOD RESEARCH PRACTICE

A Draft for Discussion

A Report of the ISPOR Health Science Committee - Task Force on Good Research Practices - Quality of Life Studies

- This working document should not be interpreted as representing the views of ISPOR or its membership.
- It is intended as a contribution to a continuing debate that ought to be informed by the widest possible constituency within the scientific community.
- The authors of this report actively welcome comments on any aspects of the document.

Task Force Chair
Paul Kind Mphil, Senior Research Fellow, Centre for Health Economics, University of York, York, United Kingdom

April 2001
**Scope, definition and purpose**

The measurement of what is generally referred to as “quality of life” in clinical trials is now a well-established practice. The term itself is not new and early references to it can be found in ancient writings but in more modern times it has assumed the status of a respectable science with its own nomenclature, customs and practice. However, its recent history only dates back some two decades or so, with published accounts detailing its antecedents being fairly well scattered throughout the scientific literature\(^1\). It is thus hardly surprising to find that there are few hard ‘facts’, proven theories or methodologies in the field of quality of life measurement as a whole, much less in its application within the field of clinical trials. The purpose of this present document is to set out some of the considerations that need to be addressed in the course of measuring quality of life in clinical trials, starting with decisions to incorporate such measures and concluding with an examination of the analysis and interpretation of data generated by their application. The pathway through this maze of issues is neither wholly plotted, nor does it lead to a single inescapable conclusion. Hence this document is intended more as an indicative guide to the topology than a step-by-step set of navigational instructions.

The paper deals with a series of issues that can arise at any point in the planning and execution of a clinical trial. Sometimes these issues have to be confronted because of the special circumstances of clinical trial design. It is possible that different solutions might be proposed were those same issues anticipated in other, non-trial settings. By the same token, there may be quality of life measurement issues that principally occur within the trial setting and for which there are no external analogues.

The structure of this report follows a logical walkway through the stages that characterise most clinical studies. There are occasional lateral forays into related areas of interest for example where the issue under consideration has implication for those with different agendas, for example where quality of life measures are used in population surveys or in economic evaluation. For the most part however, the intended pathway is that of a clinical study with particular reference to the requirements of pharmaceutical trials. Discussion is welcomed as to when and how these more

---

\(^1\) Prior to 1980 “quality of life” was not a keyword in Index Medicus and relevant material was catalogued under the term “health status”
particular requirements might be generalised to other areas in which quality of life is used to measure outcomes - for example in the evaluation of diagnostics or devices.

**Orientation, definition and context**

Treatment is almost always provided with some purpose in mind, be that the relief of symptom, the restoration of function, it almost inevitably will be to modify what would be the otherwise probable “natural” progression of disease or illness. In the past the arrival of a new treatment might have been universally welcomed as heralding a breakthrough in enabling man to combat the burden of diseases that represented a continual threat to a health. In today’s world those dramatic discoveries arrive less frequently although our expectations remain forever optimistic. The contribution of the scientist has changed somewhat too in recent times, as has their place within our world. The deferential attitudes of 19th Century society have given way to a more confrontational posture that rejects such benevolent paternalism. It is no longer accepted that expert’s views are sufficient in themselves or that the layman lacks any legitimate concern in seeking to understand the practice of science. The watchword in today’s world is “evidence”.

The clinician, health care manager and patient share common interests in ensuring that health care is provided in a timely, efficient manner that maximises benefits and, so far as is possible minimises any disbenefits, and avoids unnecessary waste. Institutional interests may accord with these objectives too. All share a concern too that treatments produce a desired effect. The Hippocratic oath (and Archie Cochrane’s equivalent\(^2\)) enjoins the physician to do no harm. Reaching a determination of whether the effects of treatment are indeed beneficial or not requires firstly that we observe the patient and their response to treatment. That process of observation is critical to the delivery of care for individual patients and forms the central rationale for the collection of data that informs evaluation in all forms of clinical study. The typical course of action in assessing “change” is to subtract standardised observations at two points in time and this report is concerned with one highly specialised form of such measurement.

\(^2\) Only provide treatment if you are able to distinguish between its benefits and its disbenefits Rock Carling Foundation Lecture 1971
The measurement of quality of life has become, in the investigation of some therapeutic areas at least, as ubiquitous an observation as that of blood pressure or weight. At any single point in time, such observations can be used to inform judgements about a patient’s health status relative to normative values for a corresponding population. In the context of the evaluation of interventions, however, it is the change in this parameter that is of particular value. Whilst quality of life measures are readily identified from the literature, there is a corresponding variety in the way that the term is defined and used. For the purposes of this report it is necessary that some attempt is made to establish an agreed description\(^3\) - a form of words that conveys the salient features of the construct that is the prime focus of the discussion that follows. Quality of life is not a new concept and early reference can be found in Greek literature. In its widest sense it embraces all aspects of our being and includes social, economic and cultural facets of our lives. Politicians recognise the virtuous nature of “quality of life” as a desired objective to be striven for. Our aesthetic welfare is enhanced through the access to music and the arts. Freedom of thought and movement, safety and security in our homes and places of work – even a place in the world of work – all these contribute to the quality of our individual lives and in aggregate to that of society as a whole. The Eisenhower report on the quality of life of the American people (Campbell, Converse and Rodgers, 1976) also noted the importance of health nested within the array of elements that comprise this global concept.

Use of the term "quality of life" invokes a number of reactions that are potentially damaging for scientists and clinicians involved in the delivery and evaluation of health care. These potential risks can escalate to unacceptable levels when associated with the promotion of new therapies. Both the lay public and health care professionals automatically accord status to the concept of "quality of life" that goes well beyond any comprehension that they might have of its definition. In fact precisely because the meaning is left unspecified it has a malleable property that renders it simultaneously valued by all and fully understood by none. Definitions of the term are to be found (refs) but since they all differ to some extent it is difficult to discern a single definition that commands universal recognition as the consensual standard. It is not unsurprising to find that the majority of studies in which ‘quality of life’ is reported make no attempt to define the applied construct – rather it is left to the reader to infer the meaning from the context in which it appears.

\(^3\) definition is too strong a term in this context
It is also common to find that need for definition is avoided by incorporating a measure that is labelled as being “a measure of quality of life” or a “standardised measure of quality of life”. It is as if the need for clarity of definition is deemed unnecessary since other researchers or the instrument developers themselves have adequately taken care of this task. Since such measures are typically the product of other researchers’ labours it would seem to be appropriate to require that end-users review available evidence for themselves before committing to the use of any measure. Clearly a review of this sort might become less pressing following increased exposure and usage. Since different measures have followed somewhat different developmental pathways and have en route, generated varying volumes of evidence of variable quality, it will always remain largely a matter of judgement as to the status of its evidential base. Whilst the definition of “quality of life” is always likely to remain problematic, it might be less challenging for the scientific community to agree upon the type of evidence that it would require before conferring that descriptive title on any measure. It is clearly not acceptable that any measure can be labelled as a “quality of life” measure on the whim of individual users. In the context of clinical trials and especially in regards to the anticipated use of data generate by such trials, a degree of consistency is demanded in scientific standards characterised by well-defined measurement using calibrated instruments. This then ought to be the goal to which we might all aspire.

In the text that follows, the preferred terminology of health-related quality of life (HrQoL) has been adopted so as to distinguish it from the more universal construct of quality of life (QoL) of which it forms a subset. In recent times a more general terminology has been introduced that places “quality of life” measurement within a more general set of patient-reported outcomes (PRO). It is not intended here to rehearse the pros and cons of this schema save to make the point that QoL or HrQoL may be observed from different perspectives and that to uniquely classify a measure under the PRO categorisation is to deny its legitimacy elsewhere. Hence, designating QoL or HrQoL as being solely within the competence of the patient raises profound difficulty when it comes to dealing with circumstances in which the patient cannot themselves undertake such assessment. How is HrQoL to be measured in patients with intellectual difficulties or where dependency issues dominate? What is the position of children or aged patients when parents or carers are asked to act on their behalf in recording HrQoL? The PRO schema is useful is
establishing a general topology in which many sorts of measurement may co-exist, but it has not solved the pre-existing dilemma as to what is really meant by the term “quality of life”.

Issues / Questions
(a) at what point in the development / application of a new measure can it be reasonably described as being a standardised measure of “quality of life”
(b) what evidence should be available and reviewed in making such a determination?
(c) Under what circumstances can the term “quality of life” be used legitimately? Should it be replaced in the context of clinical trials by “health-related quality of life” or some other term?

Justifying HrQoL measurement
In approaching the design of a new clinical trial the inclusion (or non-inclusion) of an HrQoL measure should be explicitly considered. It is by no means clear that such consideration is always undertaken, or that the basis upon which such a determination is made (when it is made) is clearly articulated and understood. It seems hardly consistent with good scientific practice to imagine that the inclusion (or exclusion) of variables within an experimental situation are not made for specific and sustainable reasons. Such decisions may sometimes appear trivial. For example, it would nonsensical not to measure blood pressure in a study of antihypertensive therapy. At a minimum, the process of formally addressing such decisions ensures that the basis on which judgements are made can be scrutinised and that the cited evidence can be revisited if necessary. The process generates both a justification and a defence. Gill and Feinstein (ref) note that such frank considerations in respect of “quality of life” are mostly absent from the clinical literature.

Clearly a default option to routinely include such measurement is as questionable as would be the position to routinely exclude. However, where current custom and practice exercises undue weight in the decision-making processes, there is a danger of adopting a semi-automatic reaction to the question as to whether HrQoL measurement should be included in a trial protocol. In practice comparability with pre-existing studies may weigh heavily in that determination but this risks perpetuating errors of judgement or measurement made by other researchers. The Karnofsky Performance Status Index (KPS) probably exceeds all other measures of its type in terms of longevity. Assuming for a moment that its performance characteristics and the level of achieved
measurement were not in doubt⁴ its selection in oncology studies owes much to the choices exercised by previous generations of researchers who repeatedly opted to include it in their studies. Breaking away from tradition and the legacy of past trial design is a costly process – so that it seems likely that KPS will continue to figure in oncology studies, despite any technical shortcomings that might be revealed by a thorough review of its credentials.

The need for replication is an ever-present determinant in current trial design and may be difficult to counter in the real world. Adoption of measures based on current fashion or an apparent new trend that might be loosely described as “me-tooism” ought to be more readily dealt with but where other trials have led the way it may be difficult to stand out for a more reasoned selection of HrQoL measure.

What then are the circumstances under which researchers should contemplate including HrQoL in trials - and as relevant, when should they not do so ? The issue has been the subject of enquiry and report by many organisations and special interest groups. Guidelines in several therapeutic areas cite the measurement of quality of life as a desirable or essential component of any clinical trial. A selection are listed in Table 1 (note : please feel free to add other examples). The question as to whether or not HrQoL should be measured in a clinical trial is one that by rights should be considered at the earliest point in the testing of a new therapeutic agent. At this point the matter is one that ought to be resolved as one of principle rather than moving directly to the selection of an appropriate measure. There are clearly situations in which HrQoL measurement is inescapable – for example in palliative care settings where disease progression is inexorable, or where relief of the burden of disease is “purchased” through treatment that levies its own toll on the patient.

---

⁴ The Karnofsky Index is at best a weighted ordinal scale
Table 1: Selected disease areas

<table>
<thead>
<tr>
<th>Disease Area</th>
<th>Management/Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer</td>
<td>Management of insulin-dependent (type 1) diabetes</td>
</tr>
<tr>
<td>Gastrointestinal disorders</td>
<td>Management of chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>Cardiovascular therapies</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Periodic limb movement disorder</td>
</tr>
<tr>
<td>Antithrombotic therapy</td>
<td></td>
</tr>
<tr>
<td>Prostate cancer screening</td>
<td></td>
</tr>
<tr>
<td>Osteoporosis</td>
<td></td>
</tr>
<tr>
<td>Bone marrow transplantation</td>
<td></td>
</tr>
<tr>
<td>Nutritional status of people with AIDS</td>
<td></td>
</tr>
<tr>
<td>Treatment of essential hypertension</td>
<td></td>
</tr>
<tr>
<td>Early discharge from the intensive care unit</td>
<td></td>
</tr>
</tbody>
</table>

The European Organisation for Research and Treatment for Cancer (EORTC) responded to the growth of clinical studies that included quality of life by publishing guidelines and procedures for all stages of the clinical trial process (Kiebart, Curran and Aaronson, 1998). In 1993, the Health Services Research Committee of the American Society of Clinical Oncology (ASCO) set up an Outcomes Working Group to consider the development of cancer treatment guidelines for use in technology assessment. Despite its terms of reference, this work took a particular view of the relative position of quality of life. Survival was deemed to be the most important outcome of cancer treatment. This presupposition goes to the heart of the justification for HrQoL measurement in that the relative value of survival and other aspects of an individual’s existence – their functional capacity and performance, their physical, social and mental welfare – are judged to capable of empirical investigation. Hence the singular character of quality of life research – namely the measurement of value.

The Heart Failure Guideline Panel sponsored by the Agency for Health Care Policy and Research (AHCPR) determined that changes in intermediate outcomes constituted insufficient evidence of the effectiveness of treatment and that the relevant outcomes were those experienced directly by patients. Health-related quality of life is central to this view. Hadorn et al recognise a dominant position for patient self-assessment of HrQoL, suggesting that clinical test results are a useful supplement to such data (Hadorn, Baker, Dracup and Pitt, 1994).
Not all conditions, treatments, mode of analysis are conducive to the measurement of HrQoL. Where the effects of ill health are short-acting or fluctuate rapidly but without significant lasting effect then it may be difficult to make a convincing case for recording HrQoL as an outcome measure. These circumstances are particularly inamicable where instruments incorporate a specific timeframe over which assessment is to be made. Clearly asking patients to assess their HrQoL over a 4 week recall period would be of little sense in a treatment that resolved symptoms within hours. Evaluation of new treatments is typically conducted with respect to some other active form of therapy so that the problems are accentuated, as it is the marginal effect on HrQoL that will be of interest. Where small differences between treatments are anticipated then in general, HrQoL measurement is unlikely to yield significant useable material in support of one or other option. This may not be a problem where equivalence is being investigated.

Even where HrQoL measurement appears justified it may be difficult to establish causal association with other variables of interest. For example, it is recognised in respiratory disease that patients with advanced emphysema not only experience severe difficulties with breathing but that their condition adversely affects their physical functioning and health-related quality of life. The patient's perception of their breathlessness is influenced by several factors that include lung function but also other, non-pulmonary factors. The causal pathway between a single clinical parameter, for example lung function, and health-related quality of life is poorly defined. In these circumstances to anticipate that "a reduction in lung volumes will improve health-related quality of life .. represents a considerable leap of faith" (Cranshaw and Evans, Lancet, 357, March 3rd 2001).

The issue of appropriateness in recommendations around practice in HrQoL measurement is seldom addressed, presumably since we lack a general standard definition. Selected statements from a recent review article in this field highlight the problem for the partially informed reader. "Performance status scales are generally considered to be insufficient as measures of HrQoL … equating changes in performance status with .. patient's quality of life is questionable" and "Developing new instruments is warranted if it is felt that existing measures do not meet predefined needs of assessment" and "it is difficult to judge the decision to utilise one instrument over another". Far from being peripheral considerations these statements reflect central issues in
the deployment of HrQoL measures in clinical trials. Some attempt at defining the conceptual basis for the measurement of HrQoL ought to be a pre-requisite for its operationalisation. Such a statement might very well conceive of HrQoL in functional terms so that a performance status scale would then be perfectly consistent with that prior specification. As to the comparison of one measure against another, this is precisely the evidence that ought to be rehearsed as part of the justification for making any choice in this area - particularly if researchers embark on the somewhat speculative business of constructing a new measure of their own.

**Issues / Questions**
(a) Should HrQoL measurement always form part of the evaluation of treatment?
(b) Under what circumstances, if any, should HrQoL not be included?
(c) How can replication errors be avoided or their consequences be minimised?

**Selecting a HrQoL measure**
If the primary task in introducing HrQoL measurement into clinical trials is to establish the justification for such measurement, then the next logical task is to consider the process of selecting a suitable instrument. As before there are few, if any certainties. In the era of evidence-based medicine it has become fashionable to turn to systematic reviews as a means of accelerating progress along the learning curve. In the field of HrQoL measurement this can be a dangerous tactic since almost all experts in the field will differ on some measurement-related issues. Review material in the field is sometimes of a very high order of quality; sometimes it is not. Instrument developers and others with significant practical experience of HrQoL measurement are often a good source of opinion, but when the advice of experts is sought, it is generally prudent to establish the biases and orientation of the oracle. It is likely, for example, that instrument developers will tend to favour their own product rather than that of other competing groups. Indeed caution should be extended to include the interpretation of some literature in the field since publication itself is no guarantee of scientific merit, particularly where there is an asymmetry in the knowledge base of authors and journal reviewers. The first rule in this area then, is to proceed with care. Ultimately there is no substitute for experience and in the best of all circumstances prospective users would be well advised to “try before they buy”.
Selection of any instrument should proceed only after a thorough appraisal of the relevant evidence bearing on the design of the instrument and its performance characteristics. For the former this means uncovering the original papers that describe the conceptualisation and development of the instrument. This may be a time consuming activity when such original material is not readily to hand. However, it would have been better to discover that an instrument planned for inclusion in a study of HrQoL in antihypertensive treatment had in fact been developed for use in assessing patients following neurosurgery. This *prima facie* mismatch was only revealed after the event when a clinical paper was published. Notwithstanding, the flawed study has been widely cited – even finding its way into advertising copy in the clinical press.

Perhaps the most commonly posed questions relating to instrument selection are linked to the expected performance of a test measure. “Is it valid?” and “Will it be sensitive enough?” might appear reasonable enough but are questions that may mask other issues. Questions concerning validity may be naïve or ill-conceived. Validity is a construct that is highly context specific – even when it is thoroughly understood. To talk of it in absolute terms is inappropriate. Validity is sometimes used as a qualifier of instrument performance, with the implication that some independent verification process has preceded the award. In point of fact, validity is itself multi-facetted and for some, a contentious construct. For example, convergence with other observations in which we have greater faith suggests the possibility of redundancy. Furthermore, some instruments have a disproportionate influence on the development of others so that “validity” established through convergence testing may be simply an artefact of a common developmental source. Association between one HrQoL measure and another may be reassuring for the end-user but might mean that both are capable of performing equally well (or badly). Construct validity tests can be informative but need to be set in the framework of a formally defined account of the descriptive basis of any HrQoL measure. Since not all HrQoL measures cover the same domains of interest, data will usually not be available with which to test the impact of any excluded domains.

The second question “Will it be sensitive enough?” is troublesome on other grounds. No trialist would deliberately include an instrument with known performance defects and at one level it is reasonable to seek to maximise the anticipated advantage promised by innovative treatment.
However, these are troubled waters. Rejecting a measure because it is expected to be “miss” the target or selecting a measure on the basis of a hoped-for “hit” suggests a technology of HrQoL measurement that is in pretty poor shape. The question could be seen as being biased towards a particular, preferred outcome whereas the nature of scientific enquiry is to leave reaching conclusions of that sort until the relevant data have been collected and analysed. A determination regarding sensitivity to change can best be made when suitable data can be analysed on a first-hand basis. This suggests that as an ideal, a proper appraisal of the performance of HrQoL measures requires early testing in the lifecycle of any new pharmaceutical. Given that this is not always possible, it is not unsurprising that researchers are forced to rely upon reported studies in the literature. This dependency is not without risk since the scientific rigour of other research can never be fully guaranteed.

An important consideration in the selection process has to be the availability of support for the end-user. Many HrQoL measures appear in the literature following a long period of development activity undertaken by a team of researchers (for example, the RAND HIS group and the EORTC). Other measures are the product of a single researcher’s creativity for which a single paper provides the only published source (for example, Karnofsky and Apgar). For those who anticipate adopting such measures there is the problem of locating appropriate evidence from the published literature, of accessing useful training material where necessary and of obtaining real-time support in implementation and data analysis. Just as with the purchase of a motor vehicle, it pays to anticipate trouble. For many HrQoL measures contact details for the principal instrument developers can be readily obtained via one of a host of internet sites or through published compendia (Spilker, 1990, McDowell, 1987). Investment of this type is rarely unproductive.

Profiles and summary indexes
HrQoL measures usually takes one of two formats – multidimensional profiles or summary indexes. These competing formats are sometimes co-existent within a single measure. For example, the appropriately named Sickness Impact Profile yields sub-scale scores for its constituent domains, but can nevertheless be represented by a single aggregate total score across all items/domains. Instrument developers may have their own reasons for adopting one or other of these formats but it is for end-users to determine the merits of the case since there is often no
compelling technical argument that tilts the balance of favour. It will be suggested that since health-related quality of life is a multidimensional construct it is inconceivable for it to be represented by a “crude” summary measure and that in any event, the process of aggregation leads to the irrevocable loss of information about its constituent elements. It is true that the profile format retains a separation of information that can be useful in describing HrQoL or in helping to understand the processes of change. However, this attribute also leaves unresolved several important questions. In particular, the contribution made by each individual dimension to overall HrQoL remains uncertain. Where the relative importance of each dimension is unknown then the end-user is free to impose their own interpretation. Whilst an improved mental health dimension score may indicate a positive contribution to improved HrQoL, how is such an improvement to be set against possible deterioration as reflected in, say, a negative change on a physical sub-scale. In the absence of any formal mechanism for aggregating such information then a self-denying ordinance on unsubstantiated claims for HrQoL would seem to be justified where profile data are invoked and where any negative change is registered.

Aggregation becomes non-negotiable where a summary index is required – as in the computation of a cost-benefit ratio. The analytic environment in which the data are to be used and the nature of the intended application will determine the precise means by which a single index measure is constructed. In a cost-effectiveness analysis it may suffice to present cost per unit change in a summary HrQoL score; for cost-utility analysis where results are presented in terms of cost/QALY ratios then clearly the summary HrQoL measure must be based upon utility weights. The origin of these weights and the means by which they have been elicited may vary according to local (national) orientation but they will almost certainly have been generated by one or other of the accepted methodologies (such as Standard Gamble or Time Trade-Off).

**Issues / Questions**

1. Under what circumstances is it necessary to aggregate multidimensional HrQoL data?
2. Can HrQoL claims be legitimately based on profile subscores and if so, then what safeguards might be needed?
3. Do summary measures of HrQoL have a useful role in clinical studies?
**Issues in analysis and reporting**

Although the impact of treatment on HrQoL may be a declared objective, clinical studies may incorporate measures that present difficulties in establishing such an effect. Investigators may have selected instruments with multiple sub-scales or indeed may have designed the study so as to include multiple instruments (for example, targeted condition-specific measures and one or more generic measures). Inevitably such a rich array of data poses special problems in its analysis, especially when compounded by repeated observations in time.

Shortcomings in the reported literature are a well-recognised phenomenon, (for example, Staquet, Berzon, Osaba and Machin, 1996) although emphasis tends to focus on technical aspects of data manipulation and analysis, for example in respect of the issue of missing data. In a more recent structured review of HrQoL papers Lee and Chi (2000) concluded that standard of reporting of HrQoL assessment and inadequate description of methodology. Development and application of structured formats for presentation of HrQoL may help to improve the standard of reporting of HrQoL in the literature. Whilst such remedies may prove useful in the short term, more entrenched factors such as the patchy review process and low levels of comprehension of technical issues are unlikely to so easily resolved.

**Missing data / compliance / consistency**

Missing data are a problem in any study, imposing real costs and inefficiencies. Variability in missing data rates can be anticipated in complex multicentre trials, especially where protocols extend over a significant time horizon, so that the costs of maintaining acceptable quality standards may be high. Where source material is open for re-inspection (for example in reading X-ray films or assaying lab tests), then it may be possible to complete partial data records. With respect to HrQoL data such options are rarely, if ever, an option. Although post hoc strategies may be constructed to partially compensate for the problem, lost data may never be fully retrieved. However, such strategies often depend upon the distribution of known responses and to some extent will always be imperfect in predicting the actual value of a missed data item.
The extent and nature of missing data will be a performance parameter of interest in the selection of any HrQoL. Assuming that, for all practical purposes, some degree of partial response is likely to be inevitable, then consideration of what constitutes an acceptable (least intolerable) level ought to influence the choice of measure. Where missing data are systematically linked to some characteristic of the respondent then there should be serious cause for concern. For example, if missing data rates arising from the use of a particular measure are consistently associated with respondent age, education or health status, then it might be that sensible to avoid using that measure as a key measure of outcome. Furthermore, where changes in health status are associated with patient non-compliance with an HrQoL measure then this would be a clear contraindication of its use. Where respondent characteristics are in doubt, for example in non-trial settings such as population surveys, consideration of the correlates of non-response becomes of increased importance. Were non-compliance to be associated with treatment side-effects or disease progression, then this too would be a cause for concern.

The literature describing the problems of missing HrQoL data is relatively rich and includes several papers with useful checklists indicating points for consideration. An example taken from one such (Fayers, Curran and Machin, 1998) is given in Table 2.

**Table 2**
Are patients with missing data items different from other patients ?
Do items comprising the scale all have similar mean values ?
Is the scale ordered or hierarchical ?
Do items comprising the scale have high correlations with each other ?
Do items comprising the scale have similar standard deviations ?
Is the item correlated with external factors or baseline variables ?

**Issues / Questions**
1. Can missing data ever be safely compensated in the observation of HrQoL ?
2. Is there a point at which non-compliance seriously compromises the viability of HrQoL data ?
3. How should strategies for estimating missing data be tested and reported ?
**Next steps**

This document has been prepared with a view to stimulating discussion of the use of HrQoL measures in clinical trials. A copy has been circulated for comment to a number of individuals and groups with known interest and competence in the field of HrQoL measurement, and has been lodged on the ISPOR website. Over the coming months it is hoped that feedback from this consultation process will provide additional material that can be incorporated in a revised version of the document that will be presented for discussion at ISPOR’s European Meeting in Cannes in September 2001 prior to a final redrafting. It is anticipated that the final version of this paper will include a practice-driven checklist designed to help improve decision-making at all levels.
Good practice in HrQoL measurement – template for a future checklist

Definition

Justification

Instrument selection

Study design

Administration

Data analysis

Reporting / claims