

This article was originally published in the
Journal of Clinical Epidemiology 2007;60:1034-1039 and is available from
<http://www.sciencedirect.com/science//journal/08954356>

Implicit self-comparisons against others can bias quality of life assessments.

Peter M Fayers^{1,2}, Anne L Langston^{3,4}, Clare Robertson³, on behalf of the PRISM trial group

¹Department of Public Health, Institute of Applied Health Sciences, University of Aberdeen, UK

²Pain and Palliation Research Group, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

³Health Services Research Unit, Institute of Applied Health Sciences, University of Aberdeen, UK

⁴Queen's Medical Research Institute, University of Edinburgh

Correspondence to:

Prof P M Fayers, Department of Public Health, University of Aberdeen, Aberdeen AB24 2TN, UK

p.fayers@abdn.ac.uk

Tel: +44 1224 559 573

Abstract

Objectives:

To explore how patient-reported health related quality of life (HRQL) and global health status are affected by use of differing personal reference frames. We hypothesised that implicit comparisons against self at an earlier time, against healthy peers or against ill patients would greatly affect patients' response values.

Study design and setting:

Patients in a randomised trial for treatment of Paget's disease completed annual HRQL questionnaires.

Supplementary questions were appended, asking the patients whether they were aware of having made implicit comparisons.

Results:

The majority of patients reported considering themselves a year ago (31% at baseline), themselves before becoming ill (23%) or other healthy people (24%), with similar proportions during follow up. Mean

HRQL scores varied substantially according to the declared frame of reference, with differences as big as

19% of the scale score, or a standardised mean effect size of 0.74 standard deviations.

Conclusion:

Reported reference frames were associated with effects of similar magnitude to the differences in HRQL that are regarded as clinically important. This may be of particular concern in trials that randomise patients to management in different settings, such as treatment at home / in hospital, or surgery / chemotherapy, and might bias or obscure HRQL differences.

Introduction

Patient-completed global questions about “overall health” and “overall quality of life” (health-related quality of life, HRQL) have been reported to be important prognostic indicators in advanced cancer, and are good predictors of survival [1,2], and similar results have been observed both for HRQL and for self-rated health in many other serious illnesses [3, 4]. Furthermore, as Fayers and Sprangers noted [5], these results are curiously robust across a wide range of seemingly vague and non-specific questions. Thus it is not at all clear what patients have in mind when they respond to such questions as “What is your overall quality of life during the past week?” Indeed, this seemingly simple question begs the query: “Compared to what?” Fayers and Sprangers speculated that patients might employ different frames of reference, resulting in responses that are derived from implicit comparisons with various peer groups or against themselves at some previous time. Comparison groups that patients may have in mind include: healthy people, such as friends or family; other patients (“Compared to all those other very ill patients I see at the clinic, I’m doing very well”); themselves prior to their illness; or themselves at some other previous time such as a year or more ago. The aims of this study were to explore whether patients use identifiable frames of reference, and the impact these might have on responses to questions about well-being or quality of life. We hypothesised that most patients would consciously have in mind one (or more) reference frames, and that these could be elicited by questioning. Identifying these implicit comparisons is important because the response levels are likely to be affected by the particular comparisons used by each patient. We postulated that ill patients who are comparing themselves against a peer group comprising healthy friends would report a relatively poor HRQL, and that those who use other patients as their reference frame would report a better HRQL.

Methods

Patients

A convenience sample was obtained from the ongoing PRISM randomised clinical trial of intensive versus symptomatic management for patients with Paget’s disease (ISRCTN 12989577). Patients were already completing annually the SF36, the EuroQoL EQ5D, and the HAQ questionnaires. For our supplementary study, three additional questions were inserted after the SF36 questionnaire. As this was an opportunistic study, some trial patients had been recruited before the additional items were included. The trial recruited 1325 eligible patients, of whom 976 completed the additional items at baseline (pre-randomisation), 1076 at one year, and 967 at two years. Severity of illness was assessed as time since the initial diagnosis, the number of bones involved, the level of deformity, and the number of fractures.

Questionnaires

The SF36 asks about “health in general” and offers response options from 1 = excellent to 5 = poor. We appended at the end of the SF36 an equivalent item, with the same response options: “How would you rate your overall quality of life during the past week?” Patients were then given an open-response question: “We realise that different people have different things in mind when they answer questions about their ‘quality of life’. What things were you thinking about when you assessed your quality of life?” Finally they were asked: “When you rated your overall quality of life, were you mainly comparing yourself against one or more of the following?” with options that included: “before you became ill”, “how you felt a year ago”, “other people with Paget’s disease”, “healthy people that you know (such as family or friends)” and “something else (please specify)”. Patients could tick one or more response options. The use of these additional questions was approved by the multicentre and local research ethics committees.

Statistical analysis

Patients were assigned to groups according to their self-reported reference frame. One-way analysis of variance (ANOVA) was used to test for differences between the groups. Effects were compared visually and plotted with 95% confidence intervals. The study sample size had been determined for the main clinical trials outcomes, and was easily large enough for this exploratory quality of life substudy. *p*-values are uncorrected for multiplicity of testing, and therefore interpreted conservatively. Regression modelling was used to explore potential explanatory factors (age, gender, age at diagnosis, and disease severity as measured by bone deformity and bone pain, and HAQ scores). Patterns of change over time were explored by considering the transition matrices: for example, 26 patients reported using healthy peers at baseline and self before ill at one year, with mean change in HRQL of 0.27, while 22 changed in the corresponding reverse direction with a mean change also of 0.27. Generalized estimating equations 5 (GEE), which can allow for correlations between successive ratings, were also used to explore patterns of change over the three time points.

Results

The mean age of the patients was 74 (range 37 to 94) and 53% were male. Approximately 14% were recruited within the first year after diagnosis, and 50% were within five years of diagnosis. Table 1 shows the distribution of the responses to the reference frame question, and the corresponding mean quality of life scores. At all time points, about 20% of patients said they had in mind how they were before they became ill, nearly a third were considering themselves a year or more previously, and about 20% were comparing themselves with healthy peers. There was also a gradual increase over time in the number of patients saying they were thinking of multiple references, reaching 19% by two years. Mean quality of life scores are also shown in Table 1. There were highly significant differences at all time points (ANOVA, $p < 0.0001$). At each occasion the ranking of the main three groups was the same, and the differences between “before ill” and “healthy peers” were statistically highly significant (all with $p < 0.0001$, pairwise *t*-tests). Figure 1 shows the means at one year, with 95% confidence intervals. The relationship between time since diagnosis and the reference frame was also explored graphically and by analysis of variance (ANOVA). There was no visible association between diagnosis time and frame of reference at either baseline or later ($p = 0.2, 0.7, 0.05$ respectively). Similarly, there was no suggestion of a relationship between reference frame and the severity of illness as measured by the number of bones involved, deformity or the number of fractures. Since the scale ranges from 1=excellent to 5=poor, the full range is 4.0. The

observed 0.75 difference between the “before ill” and “healthy peers” groups ratings at baseline corresponds to a 19% overall change in values. Differences at later times were similar (18%, 21%). The standard deviation was 1.02, giving a standardised mean effect size of 0.74 standard deviations. The self-rated health item of the SF36 showed closely similar but slightly smaller patterns (data not presented), with mean differences of 16%, 14% and 17% at the three time points. 6

Although the exact date of diagnosis was not recorded, the age in years at diagnosis was known, and from this it was estimated that 119 patients had been diagnosed in the last year. These recent-diagnosis patients had less severe disease, and reported better quality of life. In this subset of patients it might be anticipated that there would be little distinction between “before ill” and “one year ago”. There were 42 patients who reported using themselves a year ago as the reference, while for 22 it was themselves before becoming ill. In these two groups the mean quality of life scores were equal (2.91, 2.90), offering a degree of validation reassurance. Potential predictive factors, as described under methods, were explored using regression modelling and analyses of covariance, but no factors were identified as explaining either the choice of reference frame or the differences in mean HRQL scores shown in Table 1. These negative analyses are not presented here. Of the 804 patients who completed both baseline and one-year questionnaires, only 265 (33%) reported using the same reference frame on both occasions. The changes appeared to be random – for example, equal numbers of patients went from “before ill” to “healthy peers” as in the opposite direction, and similarly for other transitions roughly equal numbers went in each direction between each pair of time points. This is also reflected in the constancy of the marginal percentages at each occasion, as shown in Table 1. Regression modelling and generalized estimating equations were applied to the mean HRQL scores and also to the change-scores, but no consistent patterns of change were detected. Although the overall sample size is large, there are 7 states (reference frames) in Table 1 and thus 49 possible transition states – with relatively few patients in many of the cells.

Discussion

The hypotheses that motivated this study were substantiated by the observed results. At baseline, almost a quarter of the patients reported comparing themselves against before they were ill, in contrast to another quarter comparing themselves to healthy peers, and nearly a third said themselves a year ago; similar proportions were observed at all time points. As might be anticipated, these groups of patients returned markedly different responses for their quality of life scores and for their self-rated general health, corresponding to a large effect size of 0.74 standard deviations, or 19% on a 100-point scale. Many HRQL instruments standardise their scales to range from 0 to 100, and a difference or a change of 10 units (10%) is commonly regarded as being a large and clinically significant effect. On that basis, the differences that we observed were extremely large, and are far greater than the impacts on HRQL that are commonly observed in clinical trials. The differences observed for the self-rated health item of the SF36 showed closely similar but slightly smaller patterns; this may be a reflection of having asked patients about their reference frames for rating their overall quality of life, or it may reflect the more personal interpretation of HRQL as opposed to self-rated health. Although few patients reported using the reference frame of other patients with Paget’s disease (24 at baseline, 41 at 1 year), the HRQL values of those that did were consistent with our predictions in showing the most favourable mean scores.

The results accorded with our prior hypotheses that HRQL response patterns would be strongly affected by choice of comparator group. We contend that this implies responses do not consistently reflect the underlying HRQL, but vary according to whatever is in mind at the time of answering the questionnaire. However, a plausible alternative suggestion might be that the adaptation process leads patients to identify with a more realistic reference, namely other ill patients. We observed no evidence of consistent patterns of adaptation over time. To the contrary, only a third of the patients cited the same reference frames at successive assessments and only 13% reported the same reference on all three occasions. These changes appeared to be random. However, we have launched a qualitative study to interview patients about their thoughts, using cognitive interviewing techniques [6]. This study will also be used to test our assumption that reference frames can be elicited by questioning.

The thought processes involved in responding to HRQL questionnaires are complex and the information utilised by the participant may be objective (such as presence/absence of a confirmed medical condition) and subjective [3]. Subjective information is derived from the respondent's internal knowledge and past experience, thus producing a highly personal reference frame when rating quality of life scores. Respondents' ratings may be dependent on life experiences and contact with other people [7].

There has been limited research into the cognitive strategies adopted when patients with illness are asked to rate their (health related) quality of life. A number of small-sized qualitative studies have shown that different patients employ different reference frames, and authors have speculated about the consequences. We are not aware of other publications that have quantified the consequent effects. For example, Duggan and Dijkers [8] in a qualitative study of 40 patients with spinal cord injury noted that some with medium to low HRQL refused to compare themselves with others. One suggestion was that patients can enhance their own well-being by comparing themselves to less fortunate others. This "downward comparison" principle has long been recognised in psychology (e.g. Wills [9]; Van der Zee *et al.* [10]), but has not been evaluated and quantified for clinical HRQL scales. An editorial by Buunk *et al.* [11] provides a useful review of social comparison processes, including the effect of perceived norms, upward and downward comparison processes, and the role of cognitive biases such as unrealistic optimism. The disability paradox, in which patients typically report greater happiness and HRQL than do healthy people under similar circumstances, has been widely recognised and it has been suggested that this may be attributable to a combination of downward comparisons and response biases [12].

Social scientists working in survey research have studied the implicit use of reference frames when assessing health status. However, in a population survey context, the concern has been to remove inherent ambiguities by directing respondents to use age-standardised comparisons. Investigators have used questions such as "Compared to others of your age, how would you rate your health status?", commonly specifying relative response options (better/worse) and not absolute values (such as very bad to very good) [3,13–18]. Better average health status is reported by the elderly when an age-comparison is explicitly invoked, and studies [16,17] have also found that age-related comparisons are better predictors of survival outcome. Ubel *et al.* [18] explored whether people of different ages interpret "perfect health" in different ways, and concluded that some people recalibrate their responses according to their age. These general population studies confined their investigations to age-related comparisons of health status. HRQL

assessment in ill patients has not been investigated to the same extent as in population surveys, although we contend that the comparison issues may be even more serious if, as we observed, some patients enlist a reference group of healthy peers while others use sick patients.

Our study was not without limitations. Patients with Paget's disease are very different from those with cancer or other more common chronic diseases. The mean age was 74, and 19% were over 80. Some who ticked the box "healthy people that you know, such as friends or family" qualified their answer with a note that, given their age, they didn't have any healthy friends and that their immediate family were ill or dead. Another unanticipated problem was patients commented that they did not know and had never seen anyone else with Paget's. In oncology, for example, we would expect that the majority would meet or already know others with cancer, and that therefore a number of patients could be making comparison against others with cancer. We are, therefore, repeating this study in a bladder cancer clinical trial and in a study of breast cancer. It is hoped that these studies, together with the cognitive interviewing study mentioned above, will confirm and substantiate the results observed in these arguably atypical group of patients.

Our results imply that global questions about health and HRQL should be framed more carefully, specifying an intended reference so as to reduce the potential ambiguities. Although the age-standardised comparisons have been used in surveys, we suggest that for patients it is more important to identify comparator groups including oneself at a previous time-point, healthy (similar age) peers, and patient peer-groups.

In many randomised trials it may be expected that the reference frame utilisation will be randomly balanced across the treatment arms. Then the main consequence will be loss of efficiency because extra variability has been introduced by the ambiguity of reference frame. However, in some trials the interventions being evaluated may result in patients attending different clinics, and the varying experiences of patients may introduce serious biases. Thus in cancer trials comparing surgery against chemotherapy the patients will be exposed to very different clinical environments and may have differing experiences.

Similarly, in studies of social interventions, comparisons could be between patients treated at their own home as opposed hospital inpatients. Exposure to different peer-groups might potentially lead to serious bias. A recent Norwegian study explored the HRQL of palliative care patients who were randomised to comprehensive care at home as opposed to conventional palliative care in hospital [19]. Contrary to the investigators' prior hypothesis, no difference in HRQL was observed. In light of the present findings, perhaps the hospital-based patients were more likely to report relatively high HRQL because they were exposed to other dying patients, while those at home would have been surrounded by healthy people and might report their HRQL correspondingly lower. Thus bias cannot be ruled out, and it is possible that HRQL differences were obscured. It would have been valuable to know the reference frames used by the patients in this clinical trial.

A philosophical point about which we can only speculate is whether it is the patients' perceived HRQL that changes according to reference frame, as opposed to merely their reported HRQL. Can perceived HRQL be "improved" by encouraging a patient to make downward comparisons? This is not so dissimilar from helping patients to cope and adapt to their illness. Thus we postulate that one possible implication of the large effects that we

observed is that interventions might be directed towards influencing the choice of comparator group. Although the commonly heard admonishment to “stop feeling sorry for yourself, there are lots of people worse off than you are” is generally unhelpful, it is possible that psychotherapy could substantially alter the perception of HRQL and thereby improve a patient’s satisfaction and actual quality of life. We propose that clinical trials are needed to test such hypotheses.

Despite the inherent ambiguities and uncertainties of “overall quality of life” and “general health”, these and similar questions provide valuable indicators of survival and of treatment outcome. As a consequence of the striking results of this present study, we are now investigating other patient groups to confirm the generality of the conclusions. Further studies are planned to explore whether predictions of outcome are significantly improved by allowance for frame of reference, and to investigate the importance of routinely identifying the implicit reference group when assessing HRQL outcomes in clinical trials.

Contributors

P Fayers participated in the design of the supplementary questions, the analysis and interpretation of the results, and drafting the manuscript. A Langston participated in the design of the supplementary questions, the data monitoring, the interpretation of the results, and the drafting of the manuscript. C Robertson participated in the data monitoring, the interpretation of the results, and the drafting of the manuscript.

All authors saw and approved this submission.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

We wish to acknowledge the Arthritis Research Campaign, the National Association for the Relief of Paget’s Disease, and The Alliance for Better Bone Health (Proctor & Gamble Pharmaceuticals and Sanofi Aventis) for grant support of the PRISM trial. The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Department of Health. The views expressed are those of the authors and not necessarily those of the funding body. We also wish to acknowledge the support of the PRISM Trial Group.

References

- [1] Coates A, Porzsolt F, Osoba D. Quality of life in oncology practice: prognostic value of EORTC QLQC30 scores in patients with advanced malignancy. *Eur J Cancer* 1997; 33: 1025–30.
- [2] Osoba D. What has been learned from measuring health-related quality of life in clinical oncology? *Eur J Cancer* 1999; 35: 1565–70.
- [3] Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997; 38: 21-37.
- [4] Bjorner JB, Kristensen TS, Orth-Gomer K, Tibblin G, Sullivan M, Westerholm P. Self-rated health: a

- useful concept in research, prevention and clinical medicine. Stockholm: Forskningsrådsnämnden Report 96.9, 1996.
- [5] Fayers PM, Sprangers MAG. Understanding self-rated health. *Lancet* 2002; 359: 187-8.
- [6] Willis GB. *Cognitive Interviewing. A Tool for Improving Questionnaire Design.* Thousand Oaks: Sage Publications; 2005.
- [7] Groves R, Fultz N, Martin E. Direct questioning about comprehension in a survey setting. In: Tanur J. *Questions about Questions.* New York: Russell Sage Foundation, 1992.
- [8] Duggan CH, Dijkers M. Quality of life after spinal cord injury: A Qualitative Study. *Rehab Psychol* 2001; 46: 3-27.
- [9] Wills TA. Downward comparison principles in social psychology. *Psychol Bull* 1981; 90, 245-271.
- [10] Van der Zee KI, Buunk BP, de Ruiter JH, Tempelaar R. Social comparison and the subjective wellbeing of cancer patients. *Basic and Applied Social Psychology* 1996; 18: 453-468.
- [11] Buunk BP, Gibbons FX, Visser A. The relevance of social comparison processes for prevention and health care. *Patient Education and Counseling* 2002; 47: 1-3.
- [12] Ubel PA, Loewenstein G, Schwarz N, Smith D. Misimagining the unimaginable: the disability paradox and health care decision making. *Health Psychology* 2005; 24: S57-S62.
- [13] Eriksson I, Unden AL, Elofsson S. Self-rated health. Comparisons between three different measures. Results from a population study. *Int J Epidemiol* 2001; 30: 326-333.
- [14] Baron-Epel O, Kaplan G. General subjective health status or age-related subjective health status: does it make a difference? *Soc Sci Med* 2001; 53: 1373-1381.
- [15] Vuorisalmi M, Lintonen T, Jylhä M. Global self-rated health data from a longitudinal study predicted mortality better than comparative self-rated health in old age. *J Clin Epidemiol* 2005; 58: 680-687.
- 12
- [16] Heidrich J, Liese AD, Löwel H, Keil U. Self-rated health and its relation to all-cause and cardiovascular mortality in Southern Germany. Results from the MONICA Augsburg cohort study, 1984 – 1995. *Ann Epidemiol* 2002; 12: 338-345.
- [17] Manderbacka K, Kåreholt I, Martikainen P, Lundberg O. The effect of point of reference on the association between self-ratings health and mortality. *Soc Sci Med* 2003; 56: 1447-1452.
- [18] Ubel PA, Jankovic A, Smith D, Langa KM, Fagerlin A. What is perfect health to an 85-year-old? Evidence for scale recalibration in subjective health ratings. *Medical Care* 2005; 43: 1054-1057.
- [19] Jordhøy MS, Fayers PM, Loge JH, Ahlner-Elmqvist M, Kaasa S. Quality of life in palliative cancer

care. Results from a cluster randomized trial. *J Clin Oncol* 2001; 19: 3884-3894.
13

Reference frame	Baseline				One year				Two years			
	Patients		Quality of life		Patients		Quality of life		Patients		Quality of life	
	N	%	Mean	SD	N	%	Mean	SD	N	%	Mean	SD
Self, before ill	225	23	3.52	0.99	203	19	3.59	0.88	174	18	3.66	0.95
Self, one year ago	306	31	3.12	0.97	345	32	3.23	0.92	288	30	3.26	0.95
Healthy peers	239	24	2.77	1.09	200	19	2.87	1.07	189	19	2.83	1.00
Other Paget's patients	24	2	2.67	0.70	41	4	2.66	0.91	37	4	2.81	0.88
Something else	77	8	2.92	1.14	71	7	3.00	1.08	70	7	3.00	0.98
Multiple reference frames	84	9	3.45	0.95	187	17	3.52	1.00	185	19	3.45	1.00
No response	21	2	3.29	1.06	29	3	3.24	1.02	24	3	3.17	0.92
Total	976	100	3.13	1.05	1076	100	3.24	1.01	967	100	3.25	1.01
ANOVA p-values			< 0.0001				< 0.0001				< 0.0001	

Table 1. Mean quality of life during the past week, by reference frame.

