Quality of Life Research **12**: 523–525, 2003.

Rejoinder

Chiun Hsu[1,2], Jung-Der Wang[2,3], Grace Yao[4] & Jih-Luh Tang[1,2]
[1]*Department of Oncology,* [2]*Department of Internal Medicine, National Taiwan University Hospital,* [3]*Institute of Occupational Medicine and Industrial Hygiene, College of Public Health,* [4]*Department of Psychology, National Taiwan University, Taiwan*

We appreciate very much Dr Kind's insightful comments on our study [1]. Dr Kind points directly to the heart of debate over quality of life (QOL) research: Is there a well-established, standardized instrument that can measure all the relevant facets and dimensions of health-related QOL without bias? If such an instrument exists, can it really quantify, by an interval scale, the quality of life for our patients?

QOL measurement has been increasingly emphasized in evaluating the efficacy of cancer treatment [2]. Most studies compared the impact of different cancer treatments on QOL by reporting the summary scores of QOL questionnaires [3]. The impact on survival is either considered separately from the observed QOL scores or ignored. This may result in a biased estimation of QOL because only living patients provide the follow-up QOL data for comparison [4]. The main feature of our model is to provide a composite measure, allowing for the impact of health service on both the quantity and quality of life, by direct integration of the QOL function and the survival function of a specific cohort [5]. The integration of the QOL and survival functions has a unit of score-time such as score-month or score-year, when QOL questionnaires are used as the instrument for measurement. The results are best interpreted as the cumulative survival-weighted psychometric scores, which is the area under the curve of $Q(t)$ vs. time, for the specific QOL dimension or facet [6]. The best way to test the validity of our statistical model is to put it in the context of controled clinical trials and correlate the results of the survival-weighted psychometric scores with other objective measurement of the patients' health status.

The questionnaires we used in QOL research indeed reflect the values of the instrument developers. Interpretation and decision on the relative importance of various dimensions of such profile measures can hardly be truly disinterested [7]. Therefore, the choice of questionnaires depends for a large part on the questions the researchers wish to address. The European Organization for Research and Treatment of Cancer (EORTC)-C30 questionnaire has been shown focusing strongly on the physical functioning and clinical symptoms, which may be especially useful when the aim of the QOL study, such as that of ours, is to compare the effects of different types of treatment [8]. Cancer-specific modules that are developed by several QOL research groups may further facilitate the evaluation of disease- or treatment-related side effects.

We agree with Dr Kind's comment on the scaling issues. Several assumptions are necessary in applying the rating scales [9, 10]. The first is the equal interval of the scale descriptors. The second is the intra-personal comparability, which assumes that an individual subject will perceive in the same way the scale descriptors that define different categories on the rating scale. The third is the inter-personal comparability, which assumes that different subjects will also perceive the scale descriptors on the rating scale in the same way. The fourth is the lack of response bias, which assumes that subjects will provide response without intentional or unintentional distortion. They are quite strong assumptions, and there have been ample arguments questioning their validity [11–17].

There are two approaches trying to untangle the scaling issues. The first approach aims at enhancing the structure of the Likert-type questionnaires. It has been suggested that using Likert's points between 5 and 7 may provide a statistically better approximation to an interval scale and psychometrically distinguishable by human mind [18].

Besides, the same number of Likert points should be used in a measure to reduce the difficulty of obtaining a summation score [19, 20]. The WHO-QOL group recommended a descriptor study, which has been done in the development of the Taiwan version, to select intermediate descriptors that best correspond to the 25, 50, and 75% points between the two 'extreme' anchors for the different types of response scales [21, 22]. This careful choice of descriptors may be the closest approximation of an interval scale for use in the integration with survival on the facet level [23, 24].

The second approach uses mathematical modeling to find out the correlation of the observed item scores with the underlying scale value. Mathematical modeling, such as the Rasch model, the partial credit model, and the rating scale model, can convert ordinal scores to interval scores by nonlinear transformation under several assumptions [25–28]. Alternatively, a ranking procedure has been proposed to substitute for the rating procedure to avoid the strong assumptions of rating scales [29]. However, more studies are needed to test the clinical usefulness of these new methods in quality of life research.

Given the lack of an 'ideal' instrument for QOL research, where can we, as clinicians and researchers who care for the long-term QOL change of cancer patients, go from here? First, we may measure the patients' preference by using utility methods, such as the standard gamble. The QOL function of the study cohort obtained by the utility approach will supplement the QOL data obtained by the profile measure [30]. The utility approach has the features of an interval scale based on several assumptions [31]. Second, we may compare the QOL change between different treatments facet by facet and dimension by dimension, without a judgment of the relative merits, i.e., weighting, of the different dimensions. As we have discussed in our study, different cancer treatments may have differential impact on various QOL dimensions, which can be found more easily by a profile approach. The difference in specific dimensions shown by our approach can help in generating hypothesis rather than making conclusions; the latter can be achieved only by conducting prospective clinical trials. Third, we may conduct several cross-sectional QOL surveys for the same cohort at different time points of follow-up to improve the accuracy of estimation of the mean QOL function. This longitudinal follow-up will help fine-tune our extrapolation of QOL estimate, too. Fourth, we may correlate the results of the survival-weighted psychometric scores with other objective measurement, such as treatment-related toxicity, in the context of clinical trials. Meanwhile, we are looking eagerly forward to further refinement of the instruments used for QOL research to improve the credibility of health-related QOL measurement.

We all know that the methodology of QOL measurement is far from perfect. The development of our model marks the beginning of our continual efforts in QOL research. We agree wholeheartedly with Dr Kind that we all need to pursue high ideals and standards in our own research. The present study is not the end. It is not even the end of the beginning.

References

1. Hsu C, Wang JD, Hwang JS, et al. Survival-weighted health profile for long-term survivors of acute myelogenous leukemia. Qual Life Res 2003; 12: 503–517.
2. American Society of Clinical Oncology: Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. J Clin Oncol 1996; 14: 671–679.
3. Gotay CC, Muraoka MY. Quality of life in long-term survivors of adult-onset cancers. J Natl Cancer Inst 1998; 90: 656–667.
4. Cox DR, Fitzpatrick R, Fletcher AE, et al. Quality-of-life assessment: Can we keep it simple? J R Stat Soc Assoc 1992; 155: 353–393.
5. Hwang JS, Tsauo JY, Wang JD. Estimation of expected quality adjusted survival by cross-sectional survey. Stat Med 1996; 15: 93–102.
6. Hwang JS, Wang JD. Integrating health profile with survival for quality of life assessment. Qual Life Res 2003, in press.
7. Kaplan RM. Profile versus utility based measures of outcome for clinical trials. In: Staquet MJ, Hays RD, Fayers PM (eds), Quality of Life Assessment in Clinical Trials: Methods and Practice, New York: Oxford University Press, 1998; 69–90.
8. Kemmler G, Holzner B, Kopp M, et al. Comparison of two quality-of-life instruments for cancer patients: The functional assessment of cancer therapy-general and the European Organization for Research and Treatment of Cancer quality of life questionnaire-C30. J Clin Oncol 1999; 17: 2932–2940.
9. Brady HE. Factor and ideal point analysis for interpersonally incomplete data. Psyhcometrika 1989; 54: 181–202.

10. Yao G, Lin MR, Wang JD. A comparative study on scale descriptor selection: Heterogeneous group vs. homogeneous group. Chin J Psychol 2000; 42: 141–153. (Chinese)
11. Lam TC, Klockars AJ. Anchors point effects on the equivalence of questionnaire items. J Edu Measure 1982; 19: 317–322.
12. Duncan P. Use of extreme rating categories in ratings of child behavior. Child Study J 1989; 19: 51–64.
13. Burnaska RF, Hollmann TD. An empirical comparison of the relative effects of rater response biases on three rating scale formats. J Appl Psychol 1974; 59: 307–312.
14. Saal FE, Landy FJ. The mixed standard rating scale: An evaluation. Org Behav Hum Perform 1977; 18: 19–35.
15. Velden M, Clark C. Reduction of rating scale data by means of signal detection theory. Percept Psychophys 1979; 25: 517–518.
16. Krosnick JA, Alwin DF. A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. Public Opin Q 1988; 52: 526–538.
17. Greenleaf EA. Improving rating scale measures by detecting and correcting bias components in same response styles. J Mark Res 1992; 29: 176–188.
18. Comrey AL. Factor-analytic methods of scale development in personality and clinical psychology. J Clin Psychol 1988; 56: 754–761.
19. Yao G. How to construct a questionnaire: Using the health-related quality of life questionnaire as an example. J Occup Ther Assoc (ROC) 1998; 16: 1–24 (Chinese).
20. Yao G. Critiques on six generic health-related quality of life scales and suggestions on the further studies. Psychol Test 2000; 47: 111–138 (Chinese).
21. Bullinger M. Ensuring international equivalence of quality of life measures: Problems and approaches to solutions. In: Orley J, Kuyken W (eds), Quality of Life Assessment: International Perspectives, Heidelberg: Springer-Verlag, 1994.
22. The WHOQOL group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998; 28: 551–558.
23. Yao G, Chung CW, Yu CF, Wang JD. Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. J Formos Med Assoc 2002; 101: 342–351.
24. Lin MR, Yao KP, Hwang JS, Wang JD. Scale descriptor selection for Taiwan-version of questionnaire of World Health Organization quality of life. Chin J Pub Health (Taipei) 1999; 18: 262–270 (Chinese).
25. Rasch G. Probabilistic Models for Some Intelligent and Attainment Tests. Copenhagen: Danmarks Paedogogiske Institut, 1960/1980.
26. Wright BD, Stone MH. Best Test Design. Chicago: MESA Press, 1979.
27. Wright BD, Masters GN. Rating Scale Analysis. Chicago: MESA Press, 1982.
28. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978; 43: 561–573.
29. Yao G, Bockenholt U. Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. Br J Math Stat Psychol 1999; 52: 79–92.
30. Hays RD, Alonso J, Coons SJ. Possibilities summarizing health-related quality of life when using a profile instrument. In: Staquet MJ, Hays RD, Fayers PM (eds), Quality of Life Assessment in Clinical Trials: Methods and Practice, New York: Oxford University Press, 1998; 143–153.
31. Dolan P. The measurement of health-related quality of life for use in resource allocation decision in health care. In: Culyer AJ, Newhouse JP (eds), Handbook of Health Economics, vol. 1B. Amsterdam: Elsevier, 2000; 1724–1760.