ORIGINAL CONTRIBUTION

Choice of Response Scale for Health Measurement: Comparison of 4, 5, and 7-point Scales and Visual Analog Scale

Chisato Nagata¹, Masayo Ido¹, Hiroyuki Shimizu¹, Atsushi Misao², and Hirochika Matsuura³

To compare the feasibility of response scales with different numbers of steps for measuring health status, we assessed the ease of completing 4, 5, and 7-point scales and a visual analog scale. Four forms of the questionnaire, each of which consisted of the same ten questions about health status, but with a different type of scale among the four above mentioned, were given to 48 patients with a variety of conditions and diagnoses. The forms were attached together and the order in which they were presented was systematically varied considering permutations of forms. Respondents were asked to complete the forms following the order of the sheets. The 5-point scale form was most commonly chosen as the easiest to complete, and item omission was least frequent with this form. Similar high-reliability results in terms of Cronbach's alpha were achieved for each of the four forms. An additional survey among 153 patients confirmed the results of the first survey. The selection of the easiest scale to complete varied by sex (men preferred the 7-point scale), but was not affected by the conditions or diagnoses of the patients. The study suggests that the 5-point scale is most useful for measuring health status.

J Epidemiol, 1996 ; 6 : 192-197.

item omission, health measurement, reliability, response scale

Over the last decade, there has been increasing interest in assessing health indices. Numerous studies have been carried out to measure it in various fields¹⁻⁵. Methodological standards for developing measures of health have been advocated by several authors⁶⁻⁹. In constructing a questionnaire to evaluate health status, we need to provide a response scale for answering questions. However, few articles describing health measurement scales have explained how it was decided which response scale to use.

Jette¹⁰ reported a trial comparing the reliability of three different response modes: a multiple choice scale, a ladder scale and the use of index cards. Overall, reliability was similar for the three methods. Although this study was well designed to assess different types of response (i.e. choosing a response category, checking a rung of the ladder and sorting cards), only two response scales with different numbers of points (i.e. 4 (or 5) and 12 points) were compared. Ware and Hays¹¹ compared two methods for measuring patients' satisfaction with specific

medical encounters. One form used a 6-point response scale ranging from "very satisfied" to "very dissatisfied" and the other used a 5-point scale ranging from "excellent" to "poor". Reliability did not differ between the two methods, but the 5-point scale better predicted whether patients intended to return to the same doctor in the future, recommended the doctor to a friend, and comply with the medical regimen. However, the two methods differed not only in the number of steps, but also in the verbal description used for the response scales. When constructing continuous scales, a fundamental question needs to be addressed: how many steps should there be? Although Jaeschke et al¹² suggested that increasing response options on a scale increases responsiveness up to a point, they admitted that no one is yet sure what that point is.

Kirshner et al⁹ recommended the use of visual analog scales (VAS) or Likert scales with multiple options to evaluate the magnitude of a longitudinal change in status, but there is no available information about which scale is most appropriate.

Received August 13, 1996 ; accepted September 30, 1996.

¹Department of Public Health, Gifu University School of Medicine, Gifu, Japan.

²Misao Hospital, Gifu, Japan.

³Matsuura Hospital, Aichi, Japan.

Address for correspondence: Chisato Nagata, Department of Public Health, Gifu University School of Medicine, 40 Tsukasa-Machi Gifu, 500 Japan.

VAS are normally used to rate the overall severity of pain¹³. Cella and Perry¹⁴ reported a high correlation between the VAS and standardized measures of anxiety, depression, and distress, and supported the use of a VAS for making a rapid assessment of feeling status when a more lengthy scale is not feasible. However, there is concern that a VAS offers a confusingly wide range of choice. Hunkinsson noted that 7% of patients were unable to complete a VAS even though they could use a descriptive scale¹⁴.

To compare the feasibility of VAS and multiple choice scales with various numbers of steps for measuring health status, we conducted two surveys among patients with various conditions and diagnoses in July and August 1994. We selected several questions from the item pool prepared in our ongoing study to develop a questionnaire measuring the health of the patients with hematopoietic disorders.

MATERIALS AND METHODS

First Survey

A total of 138 questions assessing the health status of patients with hematopoietic disorders was derived from experts, including physicians and nurses in the field, and the patients themselves. This initial item pool was reduced to a set of 35 items after repeated polling among these panelists¹⁵. The items covered physical, mental, and social status. Each item asked for information about the degree of difficulty or of a problem in the daily life of the patient.

Four forms of the questionnaire using different types of response scales were provided (Forms A, B, C, and D). Each consisted of the same ten questions in the same order, randomly selected from the 35 items mentioned above (Table 1). As we were interested in continuous variables, we used continuous scales with 4, 5 and 7 points and a VAS, giving a verbal description only at each end of the scale (Figure 1). These scales have been widely used. The VAS was chosen because it is a scale with an infinite number of points between the extremes, as respondents may want to distinguish minimal differences in their health status. Rating scales with 4, 5, 7-points were used for Forms A, B, and C, respectively, and a numerical code was assigned to each step on the scale. Form D comprised a VAS anchored by the extremes of the items being measured, with 5 cm long horizontal lines. The forms were attached together, and the order in which they were presented was systematically varied considering permutations of four forms (4 x 4P4 sets of questionnaires were provided for 48 subjects). Different colored sheets were used so that the forms would be easily distinguished by the respondents.

The questionnaires were given to 48 patients treated at a general hospital. Their diagnoses were distributed across 13 diagnostic classifications and included chronic renal failure, liver diseases, and lumbar hernia. The patients were asked to

Table 1. Selected items for the questionnaire.

Question

During the past few days:

- have you felt weak?
- were you short of breath?
- were you tired?
- have you been confident of your doctor?
- were you capable of going uphill?
- were you capable of going on a trip?
- have you been worrying about illness?
- have you been worrying about treatment?
- do you have an understanding of your illness from your family?
- do you have the support of your family?

Question

e.g.) During the past few days:

- were you tired ?

4-point scale

5-point scale

7-point scale

Visual analog scale

Figure 1. Response scales for the questionnaire

answer all the questions on Forms A, B and C by encircling the code number on the scale that best represented their status during the previous few days. On Form D, they were asked to mark with a vertical slash the point on the scoring line that best represented their status during the previous few days. All forms were to be completed following the order of the sheets. On an extra page, we asked the respondents to note the color of the form they found easiest to fill in.

Second Survey

As the respondents found Forms B and C the easiest to complete, we then conducted an additional survey of 153 patients in another general hospital focusing on the response scales of these forms (i.e. the 5 and 7-point scales). The respondents were patients treated in the departments of internal medicine (26.8%), surgery (12.2%), orthopedics (18.9%), ophthalmology (37.8%) and otorhinolaryngology (4.3%).

The two forms of the questionnaire contained the same items, and the instructions for rating the items were the same as in the first survey. We also asked the respondents to express the easiness of each form as a score (from 0-100). The order of the forms was systematically varied considering permutations

of two forms.

using SAS programs¹⁷.

Data Analyses

To decide which response scale is the most feasible, we considered two factors: the ease of completing the form and the number of questions left unanswered. We compared the frequencies of these two variables for the four different forms. The effect of respondents' age and sex on their choice of response scale was also examined using chi-squared tests or Fisher's exact tests. The respondents were divided into two age groups: 55 years of age and under; and 56 years of age and over. The association between the choice of form and its presentation was also tested. As the distributions of the form presented as the first page were different between men and women in the first survey, we standardized the proportions for choice of the form in a entire group to the population in which numbers of men and women are equal and any type of form presented as the first page in the same probability (=25%) in each sex. In the similar way, the proportions for a entire group were standardized as for sex and age in the second survey. Cronbach's alpha, which is one of the indices of reliability,⁶ was calculated for each form to examine whether it varied according to the type of response scale. The codes on the scales for ten items were simply added up for scoring¹⁶. The scales were reversed on certain items so that a high score consistently represented a high health status. Spearman rank correlation coefficients among the scores of the four forms were calculated. All statistical analyses of the data were carried out

RESULTS

First Survey

Overall, 20 out of 38 respondents answered that Form B was the easiest to complete. A discernible difference in choice was demonstrated between men and women ($p < 0.05$) (Table 2): one-third of men considered that Form C was the easiest to complete, but two-thirds of women opted for Form B. The distribution of the form presented as the first page was not even in each sex (Table 3). Respondents were more likely to choose the form presented as the first page. However, Form C was less frequently presented as the first page than Form B for men, and vice versa for women.

Nine men and four women missed at least one of the 40 questions. Of these 13 respondents, seven missed items on two or more forms. The cumulative number of item omissions was greater for Form B than for Form C, but the number of respondents who missed items did not vary substantially according to the type of form (Table 4). Although the sample size was too small to clarify the characteristics of the respondents who missed questions, they were more likely to be men in the older age group (56 years of age and over) and select Form C as the easiest to complete than those who answered all questions. In total 54 items were omitted across all the forms and 43 of these were ascribed to seven men in the older age group.

Table 2. Selection of the easiest scale to complete according to sex; first survey.

Selection of the easiest scale	Males(n=30) No.(%)	Females(n=18) No.(%)	Total No.(% ^a)
Form A (4-point scale)	7 (23.3)	0 (0.0)	7 (12.3)
Form B (5-point scale)	8 (26.7)	12 (66.7)	20 (48.6)
Form C (7-point scale)	10 (33.3)	3 (16.7)	13 (23.6)
Form D (Visual analog scale)	5 (16.7)	3 (16.7)	8 (15.5)

^a Adjusted for sex and the order of form presentation

Table 3. Selection of the easiest response scale according to the form presented as the first page and sex.

Form presented as the first page	Males				Total	Females				Total
	Form selected (No.)					Form selected (No.)				
	A	B	C	D		(No.)	A	B	C	
Form A (4-point)	3	3	1	2	9	0	3	0	0	3
Form B (5-point)	2	0	5	1	8	0	3	0	1	4
Form C (7-point)	2	0	3	0	5	0	3	3	1	7
Form D (Visual analog scale)	0	5	1	2	8	0	3	0	1	4

Table 4. Numbers of incomplete responders and cumulative numbers of item omissions according to type of response scale.

Response scale	Incomplete responders	Item omission
	No.(%)	No.(%)
Form A (4-point scale)	4 (8.3)	16 (3.3)
Form B (5-point scale)	6 (12.5)	17 (3.5)
Form C (7-point scale)	4 (8.3)	5 (1.0)
Form D (Visual analog scale)	7 (14.6)	16 (3.3)

Spearman rank correlation coefficients among the scores of the four forms ranged from 0.78-0.97 (0.78 between Forms A and D, and 0.97 between Forms B and C). Similar reliability results quantitated by Cronbach's alpha were achieved with each of the four forms (0.77 for Form D and 0.80 for Forms A, B and C).

Second Survey

The proportion of the respondents who gave a higher score was larger with Form B than with Form C after adjusting for age and sex ($p < 0.05$) (Table 5). In women, a higher score was more frequently obtained with Form B than with Form C ($p < 0.05$). The finding in the first survey that men found Form C easier to complete than women did was confirmed in this second survey ($p < 0.05$). Age had little effect on choice, but younger men preferred Form C. Respondents were more likely to give higher scores with the form presented as the first page ($p < 0.05$). However, the order of form presentation did not vary by sex and age group.

The means (SD) of the scores for easiness to complete the forms were 79.8 (21.0) for Form B and 81.6 (16.5) for Form C for men, and 80.9 (16.5) for Form B and 70.2 (17.0) for Form C for women.

Item omission was noted for six men and eight women. The total number of items omitted was 24 for Form C and 22 for Form B. The number of respondents who did not answer at least one item was eight for Form B and ten for Form C. The

age and sex distributions of the respondents who missed questions were similar to those who did not miss questions. Four men and 16 women did not say which form was the easiest to complete and their distributions of age and sex were similar to those of the rest of the group. However, the proportion of item omission was significantly higher for the respondents who did not say which form was the easiest to complete.

Spearman rank correlation coefficient of the scores based on ten items was 0.96 between Forms B and C. The calculated Cronbach's alpha was 0.75 for Form B and 0.76 for Form C.

DISCUSSION

This study showed that the 5-point scale was the easiest of the four types of response scales (4, 5, and 7-point scales and a VAS) to complete when applied to a questionnaire for measuring the health. In addition, item omission was least frequent with this form. The reliability in terms of Cronbach's alpha was similar for the four types. These findings suggest that a 5-point scale should be suitable for measuring health status.

We used two factors to decide on the appropriate number of points on the scale: the respondents' choice of the easiest scale to complete and the frequency of item omission. Yamaoka et al.¹⁸⁾ have noted the distribution of response category obtained in a national survey of Japanese beliefs and philosophies. Many respondents selected the neutral middle category even if a 5-point scale was applied, and so they used a 3-point scale

Table 5. Selection of easier response scale to complete by sex and age; second survey.

Scale with higher score	Males			Females			Males&Females
	Age -55	Age 56+	Total	Age -55	Age 56+	Total	Total
	No.(%)	No.(%)	No.(% ^a)	No.(%)	No.(%)	No.(% ^a)	No.(% ^b)
Form B (5-point)	6 (28.6)	13 (37.1)	19 (32.9)	34 (60.7)	21 (51.2)	55 (56.0)	74 (44.5)
Form C (7-point)	12 (57.1)	9 (25.7)	21 (41.4)	9 (16.1)	10 (24.4)	19 (20.2)	40 (30.8)
Form B/C ^c	2 (9.5)	10 (28.6)	12 (19.0)	3 (5.4)	4 (9.8)	7 (7.6)	19 (13.3)
Unknown ^d	1 (4.8)	3 (8.6)	4 (7.1)	10 (17.9)	6 (14.6)	16 (16.2)	20 (11.5)

^a Standardized for age

^b Standardized for age and sex

^c Scores were equal between Forms B and C

^d Scores were not given

for measuring non-disease-specific quality of life. Hayashi¹⁹⁾ indicated this tendency to a neutral response was unique to the Japanese. In the present studies, however, the proportion of the respondents who chose non-central/extreme categories on the 5- and 7-point scales (i.e. category values 2 or 4 on the 5-point scale, and 2, 3, 5 or 6 on the 7-point scale) ranged from 16.3-41.2% and from 26.8- 58.8%, respectively, for all the questions. The average numbers of items for which the respondents selected non-central/extreme and central values were 3.0 and 1.9 using the 5-point scale, and 4.1 and 1.1 using the 7-point scale, respectively. The Japanese may be familiar with 5-point scale, as the evaluation system classifying pupils into five grades has been adopted in elementary and junior high schools.

We checked the frequency with which the 5-point scale was selected as the easiest to complete across the strata of different diagnoses and conditions. The choice of response scale was not affected by such variables after controlling for age and sex. Although the most appropriate type of scale might vary according to age or sex of the respondents, our findings suggest that a 5-point scale is useful for measuring health status under a variety of conditions. Therefore, we are planning to use a 5-point scale for a questionnaire measuring the health status of patients with hematopoietic disorders.

There could be a counterargument that reliability and validity of the scales should have a higher priority than respondents' preference. We wanted to assess the questionnaire with 35 items using four different response scales among patients with hematopoietic disorders. In such an assessment each respondent must provide 140 ($=35 \times 4$) separate judgements and answer additional questions to check external validity. The present study was carried out to minimize the response burden placed on those respondents. As it is likely that validity depends more on item selection than the type of response scale, the validation study was preceded by deciding upon the number of points in the scale. The study subjects were not patients with hematopoietic disorders. Although the choice of the response scale was not affected by the diagnoses of the patients in the preset study, it would be premature to reach conclusion that a 5-point scale is the most appropriate for measuring the health of patients with hematopoietic disorders. However, in the next stage, we can evaluate the reliability and validity of the questionnaire with a 5-point response scale among patients with hematopoietic disorders in a pilot study.

Although we used Cronbach's alpha as an index of reliability, the reliability based on the ten items should not be a substitute for that of the final version of the questionnaire. However, we were interested in the effect of the number of points on the scale on this index in the present study, and our findings suggest that the effect is not substantial even if more items are included from the item pool into the questionnaire.

Nishisato et al²⁰⁾ demonstrated that categorizing two variables with normal distributions reduces the correlation. They

set up distributions of two variables and rounded off the numbers to create a scale with a particular number of points. The magnitude of underestimation of the correlation using the rounded numbers increased as the number of categories decreased. The loss of correlation for seven and ten categories was quite small, but the use of five categories reduced the correlation by about 12%. Based on these results, Streiner et al⁶⁾ indicated that the correlation using the rounded numbers is equivalent to the reliability of a scale with corresponding points, if everything else is constant, and recommended that the minimum number of categories should be 5-7.

In the current study, reliability in terms of Cronbach's alpha was similar for each type of scale. We found no reduction of reliability with a decreasing number of points. In the study by Nishisato et al., rounded numbers were created from the variables that were originally set up. However, it is unlikely that respondents always select the categorized values that correspond exactly to these original values whenever the number of categories or type of scale is changed. Ikeda²¹⁾ stated empirically that the reliability coefficient tends to be reduced by increasing the number of steps, but the effect is not great because the variance of error is nearly proportional to the number of steps.

ACKNOWLEDGMENTS

This study was supported by a Research Grant for Research Committee on Epidemiology of Intractable Diseases from the Ministry of Health and Welfare of Japan.

REFERENCES

1. Evans RW, Rader M, Manninen DL, EPO Clinical Trial Group. The quality of life of hemodialysis recipients treated with recombinant human erythropoietin. *J Am Med Assoc* 1990; 263: 825-830.
2. Aaronson NK, Bakker W, Stewart AL et al. Multidimensional approach to the measurement of quality of life in lung cancer clinical trials. In ; Aaronson NK, and Backmann J, eds. *The quality of life of cancer patients*. Raven, New York, 1987: 63-82.
3. Shipper H, Clinch J, McMurray CA, Levitt M. Measuring the quality of life of cancer patients: The Functional Living Index-Cancer: Development and validation. *J Clin Oncol* 1984; 2: 472-483.
4. Wilson A, Wiklund I, Lahit T, Wahl M. A summary index for the assessment of quality of life in angina pectoris. *J Clin Epidemiol* 1991; 44: 981-988.
5. Bjordal K, Ahlner-Elmqvist M, Tolleson E et al. Development of a European organization for research and treatment of cancer (EORTC) questionnaire module to be used in quality of life assessment in head and neck cancer

- patients. *Acta Oncologica* 1994; 33: 879-885.
6. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford, Oxford, 1989.
 7. McDowell I, Newell C. Measuring Health: A guide to rating scales and questionnaires. Oxford, New York, 1987.
 8. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985; 38: 27-36.
 9. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987; 40: 171-178.
 10. Jette AM. Functional status index: reliability of a chronic disease evaluation instrument. *Arch Phys Med Rehabil* 1980; 61: 396-401.
 11. Ware JE, Hays RD. Methods for measuring patient satisfaction with specific medical encounters. *Med Care* 1988; 26: 393-402.
 12. Jaeschke R, Guyatt G. How to develop and validate a new quality of life instrument. In ; Spilker ed. *Quality of life assessments in clinical trials*. Raven, New York, 1990: 47-57.
 13. Huskisson EC. Measurement of pain. *Lancet* 1974; ii: 1127-1131.
 14. Cella DF, Perry SW. Reliability and concurrent validity of three visual-analogue mood scales. *Psycholog Rep* 1986; 59:827-833.
 15. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: Characteristics and guidelines for use. *Am J Public Health* 1984; 74: 979-983.
 16. Young FW. Scaling. *Ann Rev Psychol* 1984; 35: 55-81.
 17. SAS Institute. SAS user's guide. Ver 6.03., Inc Cary, NC, 1988.
 18. Yamaoka K, Hayashi F, Hayashi C, Misumi J, Takezawa S, Maeda M, Sato H, Oe S, Ogoshi K, QOL Research Group. A Japanese version of the questionnaire for quality of life measurement. *Ann Cancer Res Ther* 1994; 3: 45-53.
 19. Hayashi C. Statistical study on Japanese national character. *J Japan Statist Soc* 1987; Special Issue: 71-95.
 20. Nishisato S, Torii Y. Effects of categorizing continuous normal variables on product-moment correlation. *Jpn Psycholog Res* 1970; 13: 45-49.
 21. Ikeda W. Research and measurements. Tokyo: Shinyou, 1980: 89-97 (in Japanese).