Comparison between telephone and self-administration of Short Form Health Survey Questionnaire (SF-36)

María García[a] / **Izabella Rohlfs**[a,c] / **Joan Vila**[b] / **Joan Sala**[c] / **Araceli Pena**[b] / **Rafael Masiá**[c] / **Jaume Marrugat**[b,d]
and the REGICOR Investigators
[a]Institut d'Investigació Biomèdica de Girona. Hospital Universitari Dr. Josep Trueta. Girona. Spain.
[b]Lipids and Cardiovascular Epidemiology. Institut Municipal d'Investigació Mèdica (IMIM). Barcelona. Spain.
[c]Cardiology and Coronary Care Unit. Hospital Universitari Dr. Josep Trueta. Girona. Spain.
[d]School of Medicine. Universitat Autónoma de Barcelona. Barcelona. Spain.

(Estudio comparativo entre la encuesta telefónica y la autoaplicada del cuestionario de salud SF-36)

Abstract

Objective: The characteristics of the 36 item Medical Outcome Short Form Health Study Survey (SF-36) questionnaire, designed as a generic indicator of health status for the general population, allow it to be self-administered or used in personal or telephone interviews. The main objective of the study was to compare the telephone and self-administered modes of SF-36 for a population from Girona (Spain).

Methods: A randomized crossover administration of the questionnaire design was used in a cardiovascular risk factor survey. Of 385 people invited to participate in the survey, 351 agreed to do so and were randomly assigned to two orders of administration (i.e., telephone-self and self-telephone); 261 completed both questionnaires. Scores were compared between administration modes using a paired *t* test. Internal consistency and agreement between modalities were analyzed by respectively applying Chronbach's alpha and intraclass correlation coefficients. The effect of the order of administration on the test-retest difference was analyzed by one-way ANOVA for repeated measurements.

Results: Physical function, physical role and social functioning received significantly lower scores when the self-administered questionnaire was used prior to the telephone survey. When the initial survey was conducted by telephone, all Chronbach's alpha coefficients (except social functioning) scored over 0.70 in the self-administered modality. The intraclass correlation coefficient ranged from 0.41 to 0.83 for the telephone-self order and from 0.32 to 0.73 for the self-telephone order. No clinically significant effect was observed for the order of application.

Conclusions: The results of the present study suggest that the telephone-administration mode of SF-36 is equivalent to and as valid as the self-administered mode.
Key words: Research methodology. SF-36. Survey analysis. Survey research. Quality of life.

Resumen

Objetivo: El cuestionario de salud SF-36 puede ser autoaplicado o utilizado en entrevistas personales o telefónicas. El objetivo principal de este trabajo fue comparar la aplicación telefónica del cuestionario y la versión autoaplicada en una población de Girona (España).

Métodos: Diseño cruzado y aleatorizado para la aplicación de las dos formas del cuestionario. Se asignaron dos órdenes de aplicación de las encuestas (telefónica-autoaplicada y autoaplicada-telefónica). Un total de 261 personas completaron los cuestionarios. Las comparaciones entre modos de aplicación se realizaron mediante la prueba de la t de Student para datos apareados. La consistencia interna y la concordancia entre modos de aplicación se analizaron mediante los coeficientes α de Chronbach y de correlación intraclase, respectivamente. Su utilizó un modelo lineal general para medidas repetidas para evaluar el efecto del orden de la aplicación de los cuestionarios.

Resultados: Cuando se utilizó primero el cuestionario autoaplicado, las escalas de función física, rol físico y función social resultaron en una menor puntuación. Todos los coeficientes α de Chronbach fueron superiores a 0,70, excepto para la escala de función social en la modalidad autoaplicada cuando se aplicó primero la encuesta telefónica. El rango de los coeficientes de correlación intraclase fue de 0,41 a 0,83 en la modalidad telefónica-autoaplicada y de 0,32 a 0,73 en la modalidad autoaplicada-telefónica. No se observó un efecto relevante del orden de aplicación.

Conclusiones: Los resultados de este estudio indican que la aplicación de la encuesta telefónica es equivalente e tan válida como la encuesta autoaplicada.
Palabras clave: Método de investigación. SF-36. Análisis de encuestas. Calidad de vida.

Introduction

The 36-Item Medical Outcome Short Form Health Study Survey (SF-36) is a questionnaire that has been designed as a generic indicator of health status for the general population and for outpatients,

irrespective of their diagnosis. Several studies have shown it to be sensitive for detecting impairments and health problems in various outpatient settings[1,2]. This approach is also currently used to measure quality of life amongst the healthy population[3-5]. This questionnaire has been validated for use with the Spanish population as part of the International Quality of Life Assessment (IQOLA) initiative[6,7].

The questionnaire, which was initially designed to be self-administered or applied through personal or telephone interviews[3], is now also available in an electronic format[8]. When repeated over a period of time the three classic types of administration provide high degrees of internal consistency, but a relatively high degree of variation is also observed with respect to SF-36 scores obtained in some studies[9], although not in others[10,11]. These results may vary due to differences in costs and response rates[9,12]. The survey's applicability to elderly[13,14], and diverse population groups[1,15] has also been analyzed. Nonetheless, few studies have compared the level of agreement between the self- and telephone-administered modes with different designs, instruments, and results[9,12,14,16-19]. Moreover, very few studies have analyzed the two modes of administration applied to the same subjects[9,17,19].

The SF-36 questionnaire was basically designed for self- or interview-administration. In this study, the self-administered version was used in the context of a cardiovascular risk factor survey in general population because this was the least expensive and the most widely used modality. We were interested in assessing whether telephone administration mode would allow us to include quality of life information among a minimum data set for selected participants who either could not personally attend interviews at the same time as undergoing physical examinations or who failed to bring back self-administered questionnaires that had previously been mailed to them, when they attended physical examinations.

The objective of the present study was to compare the telephone administration mode and the self-administered mode of SF-36 when used with a population from Girona (Spain).

Methods

SF-36 characteristics

The SF-36 consisted of a 36-item questionnaire structured in 8 subscales and measuring several aspects of perceived health: physical functioning (10 items); role limitations due to physical problems (4 items); bodily pain (2 items); social functioning (2 items); general mental health/psychological distress and well-being (5 items); role limitations due to emotional problems (3 items); vitality; energy or fatigue (4 items); and general health perception (5 items). The number of response choices per item ranged from 2 to 6[20].

SF-36 scores ranged from 0 (lowest) to 100 (highest). A maximum score of 100 implied the absence of impairment. Ware et al[21] recommended a number of specific steps in the scoring and coding of the questionnaire. A missing value was assigned to a scale when more than half of the items were missing. Where fewer items were missing, they were replaced by the respondent's own mean score for the remaining items on the scale[20]. The 2 summary measurements of physical and mental health were also calculated.

The questionnaire was translated and validated for use in Spain, and proved to be very reliable[6]. Population reference values for this instrument are also available in Spain[22].

Design of the study and participant selection procedure

The design used for this assessment consisted of a randomised crossover administration of the SF-36 questionnaire. Some of the participants were randomly contacted by telephone and invited to take part in the trial (age range: 25-74 years). These subjects were identified by a cardiovascular risk factor survey undertaken on a random population sample from Girona, Spain, in which the overall response rate was 72%[23]. Of 385 people contacted, 351 agreed to participate in the survey (91.3%) and these were randomly allocated to 2 orders of administration (i.e., telephone-self and self-telephone). Of those who agreed to participate, 261 completed both questionnaires (response rate: 72.7% for telephone-self and 76.0% for self-telephone, respectively).

In the cardiovascular risk factor survey, self-administration was undertaken during the programmed visits, regardless of random allocation. Telephone administration of the SF-36 questionnaire was carried out by a trained interviewer 3 weeks after participants were selected in the case of self-first participants and 3 weeks prior to their visits in that of telephone-first surveys.

All participants signed their informed consent, and the confidentiality of personal data was guaranteed by the researchers.

Data on socio-demographic and clinical characteristics were collected using a standard questionnaire described in detail elsewhere[23].

Statistical analyses

Data quality was verified by checking the proportion of missing data from each scale of the questionnaire and scores were then compared between for both administration modes.

The physical and mental summary measures were also analyzed. The physical component correlates best with physical functioning, physical health, bodily pain and general health, whereas the mental component correlates best with mental health, emotional role, social functioning and vitality. With respect to the 8 SF-36 scales, it was possible to estimate scores for these summary scales with smaller confidence intervals and to reduce the number of statistical comparisons. These scales were standardized to a mean score of 50 and a standard deviation of 10[7].

The scales were described as both means and standard deviations (SD). The internal consistency and agreement between the two modes in which the questionnaire was applied were analyzed by Chronbach's alpha and intraclass correlation coefficients (ICC), respectively. The 95% confidence intervals were used to test for any significant differences in the ICC for each scale with respect to the two modes in which the questionnaire was applied. The ICC is an index of concordance for dimensional measurements with a range between 0 and 1, where ≥ 0.75 indicates excellent reliability[11].

Differences between the values obtained on each scale for the two modes of administration of the questionnaire were assessed by applying the paired *t* test at both orders of administration.

The effect of the order of administration on differences for the test-retest on each of the 8 scales and for the summary of mental and physical health obtained using the questionnaire was analyzed by one-way ANOVA for repeated measurements.

The SPSS statistical package was used for these analyses.

Results

A total of 385 consecutively examined subjects were invited to participate in the REGICOR survey of cardiovascular risk factors conducted in Girona: of these 34 declined to do so. The response rate for this survey was 72%. Of those who agreed to participate, 261 completed both questionnaires: 128 were initially assigned the telephone questionnaire and 133 were initially given the self-administered questionnaire.

There were no significant characteristic differences between participants allocated to the two orders of administration (table 1). None of the scales was missing for the telephone-administered questionnaires, either for the self-telephone or telephone-self administered mode. On the other hand, 27 scales were missing in the case of the self-administered questionnaires (10 for self-telephone and 17 for telephone-self, although this was not statistically significant). Results are shown in table 2.

The mean score for each subscale and for the two

summary measurements of physical and mental health are presented in table 3 by order and mode of administration of the questionnaire. Scores tended to be somewhat lower for the self-administered questionnaire when it was used prior to the telephone questionnaire. The observed differences reached statistically significance levels for physical function (mean [SD]) (–3.0 [14.0]), physical role (–5.0 [25.0]) and social functioning (–6.0 [22.1]). On the other hand, when the telephone was conducted first, none of the dimensions exhibited significant differences with respect to the self-administration modality.

Table 1. Characteristics of the sample of individuals participating in the study by order of administration of the questionnaire

	Telephone first, n (%)	Self-administration first, n (%)	p
Age group			
< 45	47 (42.3)	52 (44.4)	
45-64	50 (45)	43 (36.8)	
> 64	14 (12.6)	22 (18.8)	
Gender			
Female	59 (53.2)	61 (52.1)	0.87
Education			
None	5 (4.5)	4 (3.4)	
Primary school	67 (60.4)	82 (70.1)	
Secondary school	27 (24.3)	25 (21.4)	
University	12 (10.8)	6 (5.1)	0.31
Type of population[a]			
Urban	10 (9)	15 (12.8)	
Rural	101 (91)	102 (87.2)	0.35
Chronic conditions[b]	37 (33.3)	44 (37.6)	0.50
Body mass index; mean (standar derivation)	26.2 (4.0)	27.0 (5.0)	0.18

[a]Urban: people living in towns with 10,000 inhabitants or more; rural: the rest.
[b]History of at least one of the following: diabetes, hypertension, hypercholesterolemia or heart failure.

Table 2. Proportion of missing data for SF-36 scale items in test and re-test administration and by administration mode (telephone or self-administered)

	Test (n = 228)		Re-test (n = 228)	
	Telephone n = 111	Self n = 117	Telephone n = 117	Self n = 111
Physical function	0	2	0	1
Physical role	0	3	0	3
Bodily pain	0	1	0	1
General health	0	5	0	9
Vitality	0	3	0	4
Social functioning	0	2	0	1
Emotional role	0	3	0	6
Mental health	0	3	0	5
Total SF-36	0	10	0	17

Table 3. Pairs of mean scores for each dimension and standardized components for each administration modality by order of administration

	Order			Order		
	Telephone first Mean (SD)	Self second Mean (SD)	p	Self first Mean (SD)	Telephone second Mean (SD)	p
Physical function	89.1 (15.7)	87.3 (16.4)	0.101	82.6 (20.5)	85.6 (17.5)	0.015
Physical role	89.2 (29.3)	85.8 (31.3)	0.220	87.1 (30.2)	92.1 (25.3)	0.24
Body pain	80.0 (22.9)	79.8 (21.2)	0.939	77.4 (22.9)	79.9 (23.6)	0.190
General health	68.1 (19.8)	66.1 (19.4)	0.108	61.9 (19.8)	62.6 (19.8)	0.624
Vitality	70.2 (18.6)	69.3 (21.2)	0.464	66.6 (18.2)	66.1 (20.2)	0.742
Social functioning	90.4 (18.2)	90.8 (14.4)	0.792	87.3 (22.7)	93.4 (16.3)	0.002
Emotional role	93.1 (23.5)	90.3 (26.6)	0.271	88.2 (29.3)	93.3 (22.5)	0.054
Mental health	73.1 (19.0)	74.0 (19.9)	0.408	69.9 (19.1)	71.0 (19.8)	0.391
Standardized component[a]						
Physical	51.1 (7.78)	50.4 (7.14)	0.171	49.4 (7.85)	50.3 (7.21)	0.121
Mental	51.1 (8.33)	51.2 (8.54)	0.919	49.0 (10.70)	50.4 (8.97)	0.065

SD: standard deviation.
[a]The physical component correlates highest with physical functioning, physical health, bodily pain and general health, whereas the mental component correlates highest with mental health, emotional role, social functioning and vitality.

The ANOVA model with repeated measurements only showed a significant within-subject variation for physical function, physical role and social functioning in the self-telephone order of administration, with slightly higher mean scales for the telephone mode, as in the bivariate analysis. A statistically significant effect was observed with respect to order (i.e., between-groups) of administration for physical function (mean difference score: 89.1 – 82.6 = 6.55) and general health (mean difference score: 68.1 – 61.9 = 5.76).

No differences were found for the physical and mental summary scales, either for each administration modality, or for the order of administration.

Table 4 presents data relating to the internal consistency of each scale in the two orders of administration, for each mode of administration (i.e., telephone and self), and also for the degree of agreement between initial and second administration for the two orders. All Chronbach's alpha coefficients, except social functioning in the self-administered modality when initial administration was via telephone, were over 0.70. This indicated that items in each part of the SF-36 scale were internally consistent, regardless of the mode of administration.

Agreement between the two modalities of administration, as measured by the ICC, ranged from 0.41 to 0.83 for the telephone-self order and from 0.32 to 0.73 for the self-telephone order of administration. Only vitality showed higher values for the telephone-self order according to 95% confidence intervals of the ICC (table 4). Physical role, social functioning, and emotional role all scored lower than on the rest of the scales, regardless of the mode of administration.

Table 4. Intraclass correlation coefficient (ICC) between telephone and self-administration of the SF-36 questionnaire and Chronbach's alpha for each order of administration

			Chronbach's alpha	
	ICC	95% CI	1st administration	2nd administration
Telephone/self-administration				
Physical function	0.68	(0.58-0.77)	0.84	0.84
Physical role	0.48	(0.34-0.61)	0.96	0.92
Body pain	0.60	(0.48-0.70)	0.83	0.82
General health	0.77	(0.68-0.83)	0.74	0.76
Vitality	0.76	(0.68-0.83)	0.73	0.82
Social functioning	0.48	(0.33-0.60)	0.77	0.50
Emotional role	0.41	(0.25-0.54)	0.90	0.87
Mental health	0.83	(0.76-0.88)	0.80	0.84
Standardized physical component	0.73	(0.62-0.80)		
Standardized mental component	0.72	(0.62-0.80)		
Self administration/telephone				
Physical function	0.72	(0.62-0.79)	0.88	0.83
Physical role	0.59	(0.46-0.69)	0.92	0.95
Body pain	0.53	(0.39-0.64)	0.82	0.84
General health	0.67	(0.57-0.76)	0.75	0.73
Vitality	0.56	(0.42-0.66)	0.73	0.74
Social functioning	0.35	(0.18-0.49)	0.74	0.71
Emotional role	0.32	(0.17-0.47)	0.89	0.88
Mental health	0.73	(0.64-0.80)	0.78	0.78
Standardized physical component	0.69	(0.58-0.77)		
Standardized mental component	0.64	(0.52-0.73)		

Discussion

In this study, the telephone administration of the SF-36 quality-of-life questionnaire proved to be equivalent to the self-administered mode, with no significant differences being found in comparisons between the two methods. Self-administration of the SF-36 questionnaire is relatively cheap and reliable and requires minimal resources. However, better understanding and response rates are obtained through telephone administration[9,16-18]. Furthermore, telephone administration requires less time[14], and provides fewer missing items[9,12,16-18]: it also tends to provide more favourable health ratings[12,16]. However, it should also be noted that in some studies telephone administration has produced the least favourable results[9]. The degree of agreement between these two modes of survey administration is still a matter for some controversy and previous studies comparing the two modalities have differed quite significantly with regard to both design and the type of questionnaire administered (table 5).

In our study, no systematic differences in health rating were found between orders, except for more favourable ratings for the telephone mode when it was administered after the self-administered modality and even then, this only occurred at 3 of the 8 scales: physical function, physical role and social functioning. These differences for physical function and physical role may not be important, as they were not clinically significant[14].

The discrepancy was greatest for social functioning, which was one of the most variable items. This result was consistent with findings from other studies[14]. Moreover, it is worth noting that since a multiplicity of paired *t* tests was required for this study, some of the apparently significant findings may have been the product of chance.

Another important finding was that the telephone mode produced no missing answers for any items, regardless of the order of administration. In contrast, the self-administered mode produced missing items for both administration orders: this confirmed other findings cited in the literature[9,12,16-8].

The effect of the order of administration was negligible and, by and large, non-statistically significant, although statistically significant differences were observed with the self-telephone administration for physical function and general health. This interpretation is based on the fact that differences of > 7 points in the physical domains and of > 10 points in the mental domains are regarded as substantial and clinically meaningful[14,24]. Moreover, mode effects are apparent when subjects systematically respond differently because of the mode itself, or when the response received is a function of the mode concerned[18].

In the telephone-self administration order, the ICC ranged from moderate (0.5-0.7) to high (> 0.7), except for physical role, social functioning, and emotional role. In the self administration-telephone order, the ICC was also moderate-to-high, except for social functioning and emotional role. The ICC for physical and emotional roles

Table 5. Characteristics and results of studies comparing survey administration modes					
Study	Instrument	Administration mode	Population	Period between administration modes	Conclusions
Duncan, 2005[18]	Stroke Impact Scale	Telephone/mail. Randomized	Veteran stroke population	1 week	Good test-retest reliability but, overall, the telephone presented better agreement
Aitken, 2004[17]	Skin cancer questionnaire	Telephone/mail. Randomized crossover	Community based	1 month	Agreement between methods was similar regardless of whether the interview or the questionnaire was administered first
Hawthorne, 2003[19]	Assessment of Quality of Life	Telephone/telephone-filtered/ mail. Randomized	Community based	2 weeks	Mail and telephone interview were directly comparable
Perkins, 1998[16]	SF-36	Telephone/mail. Randomized	Community based	Data not available	Significant differences were found between modes in internal consistency reliability estimates. Health rates were more favourable for telephone administration
Weinberger, 1996[9]	SF-36	Telephone/face-to-face/mail. Randomized crossover	Patients with chronic medical conditions	3 days/14-30 days	Scales were internally consistent for all 8 scales, regardless of mode of administration. Relative large variation in SF-36 over time was found
Weinberger, 1994[14]	SF-36	Telephone/mail. Not randomized	Elderly patients prescribed multiple medications	1 month	The two modes of administration may not produce interchangeable results

437

and for social functioning could be considered low for both orders of administration. One possible explanation for this could be the large intra-individual (temporal) variation, which has led some authors to conclude that SF-36 is not acceptable for use in research or clinical practice[9,14]. However, other studies[10,11] found no significant variation in SF-36 scores when surveys were repeated over longer periods. Our results agree with the latter studies for the majority of scales of the questionnaire, regardless of the mode of administration. Moreover, low correlations for scales related to social health, and more specifically to emotional role and social functioning, have been described in other questionnaires. This is due to the fact that the items on these scales possibly reflect different aspects of social health and the problems associated with the instrumentalisation of the emotional domain[6,9,11,18,25].

As proposed by Stewart and Ware[26], reliability is acceptable for group comparisons and for measuring functioning and well-being when the Chronbach's alpha exceeds 0.7. In our study, all scales (except for social functioning in the self-administration of the telephone/self-administration mode), had Chronbach's alpha scores of greater than 0.72, regardless of the randomization group. This high internal consistency was consistent with that observed in similar studies and supports the utility of the SF-36 in health service research, although some authors noticed that the relatively large variation in SF-36 scores over very short intervals, which was not the case in our study[11], may reduce its usefulness as an evaluative instrument[9,14].

The use of the summary scales provided a low degree of variability and greater degree of reliability. However, their use depends on which differences between groups at any given point in time, or which changes in health status over a period of time are of interest. If the evaluation relates to a particular SF-36 scale, a summary measure is less likely to capture differences other than those analysed by that particular scale.

On the other hand, insofar as differences in health are comparable across the most inter-correlated SF-36 scales, summary measures may prove more useful than any of the specific scales themselves[7].

Other studies into the mode of administration of health-related quality of life questionnaires have reported contradictory results. Wu et al[27] suggested that few differences can be observed between self- and interview-administration in the case of brief questionnaires relating to health-related quality of life, and that the resulting data could be pooled and analyzed together. Meanwhile, Lyons et al[28] concluded that the personal interview exaggerates health status with respect to the self-administered modality. More importantly, despite its greater cost, the face-to-face interview mode is more valid than the self-administered questionnaire, particularly as the latter may not be so reliable for lower-income pa-

tients and/or those belonging to ethnic minorities[29].

Cost analysis was beyond the scope of the present analysis, although the average cost of the mail or self-administered mode was half that of the average cost associated with the telephone surveys carried out in other studies[16-18].

The results of the present study suggest that the telephone administration of the SF-36 questionnaire is equivalent to, and as valid as, the usual self-administered mode.

The telephone administration mode would allow us to include quality of life information among a minimum data set obtained from selected participants who could not attend personal interviews. Those who did not bring the self-administered questionnaires that had been sent to them beforehand, together with their appointment details, could be asked to complete the required information for at the time of their physical examinations in the context of a survey. Furthermore, in situations in which response rates for self-administered questionnaires are lower than those for telephone surveys, the higher cost of the telephone surveys may be justified[17].

In conclusion, the choice of how to administer the survey should not be made on the basis of costs alone, as there are other issues that are relevant to data quality that relate to the mode of administration itself.

References

1. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Test of data quality, scaling assumptions, and reliability across diverse groups of patients. Med Care. 1994;32:40-66.
2. Ruta DA, Garrat AM, Leng M, Russell IT, MacDonald LM. A new approach to the measurement of quality of life. The Patient Generated Index. Med Care. 1994;32:1109-26.

3. Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ. 1992;305:160-4.

4. Perneger TV, Leplège A, Etter JF, Rougemont A. Validation of French-language version of the MOS 36-item short form health survey (SF-36) in young healthy adults. J Clin Epidemiol. 1995;48:1051-60.

5. Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol. 1998;51:903-12.

6. Alonso J, Prieto L, Antó JM. La versión española del SF-36 Health Survey (Cuestionario de Salud SF-36): un instrumento para la medida de los resultados clínicos. Med Clin (Barc). 1995;105:771-6.

7. Ware JE, Gandek B, Kosinski M, Aaronson N, Apolone G, Brazier J. The equivalence of the SF-36 Summary Health Scores estimated using standard and country-specific algorithms in 10 countries: Results from the IQOLA Project. J Clin Epidemiol. 1998;51:1167-70.

8. Ryan J, Corry J, Attewell R, Smithson MJ. A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. Qual Life Res. 2002; 11:19-26.

9. Weinberger M, Oddone EZ, Samsa GP, Landsman P. Are health-related quality-of-life measures affected by the mode of administration? J Clin Epidemiol. 1996;49:135-40.

10. Heyland DK, Hopman W, Coo H, Tranmer J, McColl MA. Long-term health-related quality of life in survivors of sepsis. Short Form 36: a valid and reliable measure of health-related quality of life. Crit Care Med. 2000;28:3599-605.

11. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol. 2003;56:730-5.

12. McHorney CA, Kosinski M, Ware JE Jr. Comparisons of the costs and quality of norms for the SF-36 survey collected by mail versus telephone interview: results from a national survey. Med Care. 1994;32:551-67.

13. Martikainen P, Aromaa A, Heliovaara M, Klaukka T, Knekt P, Maatela J. Reliability of perceived health by sex and age. Soc Sci Med. 1999;48:1117-22.

14. Weinberger M, Nagle B, Hanlon J, Samsa G, Schmader K, Landsman P. Assessing health-related quality of life in elderly patients: telephone versus face-to-face administration. JAGS. 1994;42:1295-9.

15. Jenkinson C, Wright L, Coulter A. Criterion validity and reliability of the SF-36 in a population sample. Qual Life Res. 1994;3:7-12.

16. Perkins JJ, Sanson-Fisher RW. An examination of self- and telephone-administered modes of administration for the Australian SF-36. J Clin Epidemiol. 1998;51:969-73.

17. Aitken J, Youl P, Janda M, Elwood M, Ring I, Lowe J. Comparability of skin screening histories obtained by telephone interviews and mailed questionnaires: a randomized crossover study. Am J Epidemiol. 2004;160:598-604.

18. Duncan P, Rever D, Kwon S, Lai SM, Studenski S, Perera S, et al. Measuring stroke impact with the stroke impact scale. Telephone versus mail administration in veterans with stroke. Med Care. 2005;43:507-15.

19. Hawthorne G. The effect of different methods of collecting data: mail, telephone and filter data collection issues in utility measurement. Qual Life Res. 2003;12:1081-8.

20. McDowell I, Newell C. Measuring health. A guide to rating scales and questionnaires. 2nd. ed. New York: Oxford University Press; 1996.

21. Ware JE. SF-36 Health survey. Manual and interpretation guide. Boston: The Health Institute; 1993.

22. Alonso J, Regidor E, Barrio G, Prieto L, Rodríguez C, de la Fuente L. Valores poblacionales de referencia de la versión española del Cuestionario de Salud SF-36. Med Clin (Barc). 1998;111:410-6.

23. Masiá R, Pena A, Marrugat J, Sala J, Vila J, Pavesi M, and the REGICOR Investigators. High prevalence of cardiovascular risk factors in Gerona, Spain, a province with low myocardial infarction incidence. J Epidemiol Community Health. 1998;52:707-15.

24. Guyat GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med. 1993;118:622-9.

25. Badia X, Alonso J, Brosa M, Lock P. Reliability of Spanish version of the Nottingham Health Profile in patients with stable end-stage renal disease. Soc Sci Med. 1994;38:153-8.

26. Stewart AI, Ware JE, editors. Measuring functioning and well-being. The medical outcomes study approach. Durham: Duke University Press; 1993.

27. Wu AW, Jacobson DL, Berzon RA, Revicki DA, Van der Host C, Fichtenbaum CJ. The effect of mode of administration on medical outcomes study ratings and EuroQol scores in AIDS. Qual Life Res. 1997;6:3-10.

28. Lyons RA, Wareham K, Lucas M, Price D, Williams J, Hutchings HA. SF-36 scores vary by method of administration: implications for study design. J Public Health Med. 1999;21: 41-5.

29. Sullivan LM, Dukes KA, Harris L, Dittens RS, Greenfield S, Kaplan SH. A comparison of various methods of collecting self-reported health outcomes data among low-income and minority patients. Med Care. 1995;33(Suppl):AS183-94.

439