

Validation of quality of life and functional measures for older people for telephone administration

Ssu-Yu (Suei) Lin BHB; **Ngaire Kerse** FRNZCGP, PhD, MBChB; **Christine McLean** Dip Ind. Relns; Dip Couns Theory; **Simon A Moyes** BSc, MSc

ABSTRACT

INTRODUCTION: Quality of life (QoL) and functional status are important aspects of health especially for older people. Efficient and valid ways of measuring older people's health is of great importance.

AIM: This project aims to establish the reliability of use of (1) a quality of life measure, the WHOQOL-BREF, and (2) a functional measure, the Nottingham Extended Activities of Daily Living scale (NEADL), for use over the telephone.

METHODS: With ethical approval, patients over age 75 years (65 years if Maori) have been enrolled in the BRIGHT trial; a randomised controlled trial testing a practice-based screening initiative to prevent disability. Participants with possible disability, defined as being unable to get in and out of the car or take hot drinks from one room to another, completed a telephone interview including the two measures. Seventy participants repeated the two measures during a face-to-face interview within three months.

RESULTS: Both WHOQOL-BREF and NEADL scores for the two forms of administration produced high Pearson correlation coefficients. There was good agreement for the WHOQOL-BREF as shown by the Bland-Altman graphs; however there was a tendency of a greater negative difference the greater the average score became (higher level of function) for the NEADL.

DISCUSSION: This study shows that telephone interviews can generally provide a valid method to assess the quality of life and function in older people.

KEYWORDS: Reliability; Quality of life; WHOQOL-BREF; functional status; NEADL; telephone administration

Department of General Practice and Primary Health Care, School of Population Health, The University of Auckland, Auckland, New Zealand

Introduction

The ageing of the population in New Zealand (NZ) means that by 2051 over a quarter of the population will be 65 years old and over, similar to many OECD countries.¹ This demographic change will increase demands on the health systems delivering interventions to improve and maintain quality of life and physical function. A valid and efficient measurement of quality of life (QoL) and function in older people is of great importance.

Health practitioners are increasingly recognising that the measurement of diseases and pathol-

ogy alone are not sufficient determinants of health status.² Quality of life measures reflect an individual's multi-dimensional well-being and are often not well understood.³ Many instruments that currently measure quality of life do not take into account several factors that are considered fundamental to the health of older people.³ The World Health Organization (WHO) defines quality of life as 'an individual's perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns'.⁴ This concept reflects the view that quality of life is a subjective perspective.

J PRIMARY HEALTH CARE
2010;2(1):35-42.

CORRESPONDENCE TO:

Ngaire Kerse
Professor, Department of General Practice and Primary Health Care, School of Population Health, The University of Auckland, PB 92019, Auckland 1001, New Zealand
n.kerse@auckland.ac.nz

The WHOQOL-BREF is a condensed version of the WHOQOL-100 and was developed by the WHO to incorporate all dimensions associated with quality of life.⁵ While results from trials examining the WHOQOL-100 indicate that this instrument may not be comprehensive enough, it is, however, appropriate and relevant for measuring quality of life in older people.⁶ The WHOQOL-BREF asks about satisfaction and ability and yields domain scores for physical, psychological, social relations and environment related quality of life and therefore provides an appropriate and efficient way to measure quality of life in the older population. Quality of life instruments provide a means to measure how beneficial health interventions are as opposed to only improving a set of symptoms.⁷ This is extremely significant for the older population as this takes into account that the benefits of life experiences and comfort may outweigh the need to increase the lifespan or suppress symptoms.

Instruments that measure the level of function give an indication of individual needs and health services required for independent living. Because of the accumulation of comorbidities and frailty, older people are higher users of health services.⁸ It has been recognised that older people may have lower levels of physical function than the general population. The Nottingham Extended Activities of Daily Living (NEADL) has been shown to be a valid and reliable indicator of physical disability in older people⁹ and is an accurate predictor of the ability of the individual to remain independent and safe in a community.¹⁰

There are many assessment instruments that can be used to evaluate the health status and requirements for older people; however, a need still exists to supply this information in an economical, time efficient and valid manner. Previous trials have shown that interviewer-administrated measures have a more positive and socially desirable response than self-completed surveys,¹¹ suggesting that the mode of administration can greatly influence results. A face-to-face interview is more time consuming and expensive to administer, especially when there are large numbers of people to be assessed. A more efficient form of interviewer-administration is telephone-administration. If quality of life and functional measures

were valid and reliable over the telephone, cost of health services assessments and research costs would be able to be optimised.

The aim of this research is to assess the reliability of (1) a quality of life measure, the WHOQOL-BREF, and (2) a functional measure, the NEADL for use over the telephone.

Methods

Instruments: WHOQOL-BREF and NEADL development and scoring

The WHOQOL-BREF is a condensed 26-item version of the WHOQOL-100, which was developed to assess the quality of life by the WHOQOL Group.⁵ WHOQOL-BREF consists of one item from each of the 24 facets of Quality of life from the WHOQOL-100 in addition to two 'benchmark' items from the general facet on general health and overall QoL.² Therefore four domains of QoL were created; physical health, psychological, social relations and environmental health.

Two typical questions:

1. *'Do you have enough energy for everyday life?'*
Response categories include: *Not at all, A little, Moderately, Mostly, Completely* yielding a score of 1-5.
2. *'How satisfied are you with your ability to perform your daily living activities?'*
Response categories include *Very dissatisfied, Dissatisfied, Neither satisfied nor dissatisfied, Satisfied, Very satisfied*, yielding a score of 1-5.

The WHOQOL-BREF was scored based on four domains: physical, psychological, social relations and environment. All items were rated on a 5-point scale with a higher score indicating a better quality of life. The domain scores were calculated by multiplying the mean of all items included within the domain by four.

The Nottingham Extended Activities of Daily Living (NEADL) provides a functional measure, which is a measure of physical disability and independence.⁹ The NEADL asks whether the older person 'does' a range of activities *'on their own, on their own with difficulty, with help, or is*

unable to do’ them. The activities include four domains: mobility, in the kitchen, domestic tasks and leisure activities comprised of 22-items of activities of daily living.

The NEADL was scored based on how easy or difficult participants felt in performing extended activities of daily living. A score of 1 meant participants were able to perform the activity on their own or on their own with difficulty. A score of zero meant that participants were unable to perform the activity or required help. The total NEADL score was calculated by adding up all individual scores, with a lower total score indicating a lower function.

Participants

This study is part of a large, randomised controlled trial—the BRIGHT Trial. Testing a practice-based screening initiative to prevent disability in the older population, a birthday card containing a short questionnaire in it was sent by the practice to the intervention group participants on their birthdays. The intervention began approximately

WHAT GAP THIS FILLS

What we already know: The WHOQOL-BREF and the Nottingham Extended Activities of Daily Living scale (NEADL) have been shown to be appropriate and relevant to measure quality of life and function in older people. Previous trials have investigated the measures of health-related quality of life and function through face-to-face interviews; however few studies have examined the use of a more efficient and cost-effective approach.

What this study adds: This research paper validates the telephone administration of these instruments in older people, hence providing a more efficient approach to measurement of these important issues. As older people are high users of health services, this paper has implications for clinicians, health evaluators and researchers.

by personal letter from their general practitioner. Written informed consent was established by mail after a telephone conversation to explain the study.

Procedure: administration of instruments

All the participants in the BRIGHT Trial were interviewed over the telephone to establish quality of life (WHOQOL-BREF) and function

Participants with disability were defined by a positive response to either or both of two items from the NEADL instrument; inability to get in and out of the car or unable to take hot drinks from one room to another without assistance

three to six months after data presented here was collected. With ethical approval, people aged 75 years (65 years for Maori) or older were recruited from general practices in three NZ District Health Board regions.

Eligibility criteria included living in the community—which includes retirement villages—aged 75 years or more (65 years or more for Maori). Those who live in residential care, were terminally ill or are not able to communicate in English were not eligible. One interviewer was fluent in Te Reo for Maori participants. The eligible participants were identified from the list of enrolled patients at interested practices and invited to participate

(NEADL), as a baseline measure for the outcomes for the trial. Participants with disability were defined by a positive response to either or both of two items from the NEADL instrument; inability to get in and out of the car or unable to take hot drinks from one room to another without assistance. This was considered a screen for disability to identify those who would go on to complete a more detailed physical assessment. This subset of all the participants completed their telephone interview and, at a later date, received a face-to-face interview in which the two measures, WHOQOL-BREF and NEADL, were completed along with a more detailed physical assessment (reported elsewhere). The in-person interview was

conducted within three months of the telephone interview by trained interviewers who were unaware of the participants' previous telephone interview scores. Only those completing the interviews within three months of each other were included.

Statistical analysis

Kappa statistics for the NEADL and kappa statistics using the Fleiss-Cohen weighting for the WHOQOL-BREF were calculated for each item and presented as the proportion of high correlation and proportion of modest correlation items. The percent of cases agreeing on each item score is reported. Scores from the telephone interviews were compared with scores from the face-to-face interviews. Pearson correlation coefficients and intra-class correlation coefficients (ICC) of the two scores on the four WHOQOL-BREF domains from telephone and face-to-face interviews

and the total NEADL scores, telephone and face-to-face, were calculated to determine the relationship between the two administration methods. The limits of agreement between the administration styles were calculated and visualised by the Bland-Altman graphs,¹² which were constructed by plotting the score differences [score difference = telephone score – face-to-face score] against the score averages [averages = (telephone score + face-to-face score)/2]. These graphs illustrate the limits within which 95% of the individual participants' score differences would be expected to lie.¹²

Results

Participant characteristics

Three thousand eight hundred and ninety-two patients participated in the BRIGHT Trial, of whom 98 (2.5%) were identified as possibly disabled using the NEADL screen over the telephone. Of these 98, 28 had completed the face-to-face interview more than three months after the telephone interview and were excluded as their status was likely to have changed. The remaining 70 completed the two assessments within three months. This study reports findings for this subsample of 70 only and their characteristics are reported in Table 1. The minimum time between the two interviews was eight days and four persons were re-interviewed in less than 14 days.

Kappa scores, mean differences and correlation coefficients

Kappa scores for each item on the NEADL (22 items) and the WHOQOL-BREF (26 items) showed modest reliability. For the NEADL, 10 items (45%) had a kappa score greater than 0.6, five (23%) between 0.6 and 0.4 with the remainder being of modest kappa ($n =$ seven items). Examining the lowest kappa score items, 'Do you manage to make yourself a hot drink?' identified 67 in agreement on 'yes' with two participants saying 'yes' on the phone and 'no' face-to-face and one saying 'no' on the phone and 'yes' face-to-face. Exact matches of score occurred in over 80% of cases for 20 of the 22 items. For the WHOQOL-BREF the kappa scores on individual items were greater than 0.4 for 13 items with the remainder being less than

Table 1. Sociodemographic characteristics of the participants

	N/Mean*	%/(SD)†
Number of participants	70	
Age, years	79.1	(5.4)
Male	32	45.7%
NZ European	46	68.7%
Other European	12	17.9%
NZ Maori	7	10.5%
Married	39	56.5%
Single	3	4.4%
Divorced	4	5.8%
Widowed	23	33.3%
Live alone	30	42.9%
Live with spouse	39	55.7%
Live with family	1	1.4%
From Canterbury	46	65.7%
From Wellington	24	34.3%
Number of prescription medications	5.0	(3.7)
Number of over the counter medications	0.5	(1.1)
Systolic Blood Pressure, mmHg	146.7	(24.6)
Diastolic Blood Pressure, mmHg	82.8	(13.0)

* Number of participants or the mean value.

† Percentage of participants or the standard deviation.

Table 2. The Mean Scores for the two forms of administration of the WHOQOL-BREF and NEADL assessments

Domain	WHOQOL-BREF				NEADL
	Physical health	Psychological	Social relationships	Environment	Total
N	70	70	70	70	70
Telephone	15.02	15.54	16.46	16.35	17.33
Face-to-face	15.26	15.63	16.71	16.63	17.19
Mean difference (SD)	-0.237, (1.98)	-0.091, (1.391)	-0.257, (1.842)	-0.284, (1.423)	0.1429, (2.155)
p-value of difference	0.32	0.58	0.25	0.10	0.58
Correlation Coefficient*	0.762	0.744	0.673	0.699	0.851
ICC (95% CI)	0.76 (0.64, 0.84)	0.74 (0.67, 0.83)	0.67 (0.57, 0.77)	0.69 (0.60, 0.77)	0.82 (0.73, 0.89)

WHO = World Health Organization

WHOQOL-BREF = Brief quality of life instrument

NEADL = Nottingham Extended Activity of Daily Living scale

* Pearson correlation coefficient.

SD = standard deviation

ICC = intra-class correlation

0.4 (50%). Exact matches were seen in over 50% for 20 of the 26 items.

The mean difference in scores for the face-to-face and telephone interviews of the WHOQOL-BREF and NEADL assessments are shown in Table 2. Table 2 shows the Pearson correlation coefficients calculated for the two assessment instruments. All four domains of the WHOQOL-BREF produced correlation coefficients greater than 0.70 which indicates a high degree of correlation between telephone and face-to-face interview administration. The correlation coefficient calculated for the Total NEADL scores is above 0.80 showing a high level of correlation between the two different forms of interview administration. (See Figures 1–5, correlations shown and Table 2, correlations reported.)

Bland-Altman graph results

The mean differences between the two scores are graphed against the average of the two scores in the Bland-Altman graphs (Figures 1 and 2). Mean differences were small and largely fell within the limits of agreement (lines 2 SD from the mean difference). The closer the individual plots are to

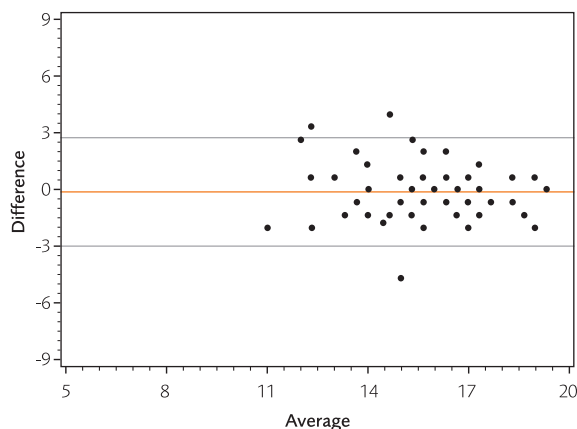
the horizontal 0 axis, the greater the overall level of agreement between the two forms of interview administration. The Bland-Altman graphs of the four WHOQOL-BREF domains are illustrated in Figures 1a to 1d. Figure 1a shows the *Physical Health Domain* which has a range of -5.7, 6.3 out of a total of 20 indicating that there is good agreement and no bias according to the mode of administration as the individual plots are fairly evenly spread above and below the horizontal line. The *Psychological Domain* is illustrated in Figure 1b and Figure 1c illustrates the *Social Relationships Domain*. The *Environment* is the last domain of the WHOQOL-BREF and is shown in Figure 1d. Lastly, the Bland-Altman graph for the Total NEADL is shown in Figure 2.

Discussion

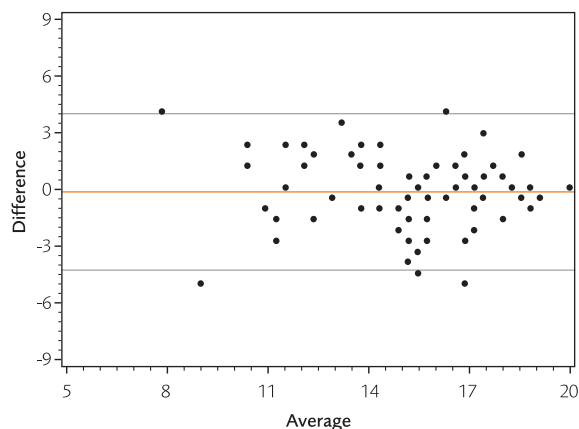
The present study is one of the first that compares telephone with face-to-face administered interviews for both a quality of life and functional measure in older people. This study demonstrates acceptable agreement between the telephone interview scores and face-to-face interview scores for the WHOQOL-BREF quality of life and NEADL function instruments. While

Figure 1. Bland–Altman Graph for the WHOQOL-BREF

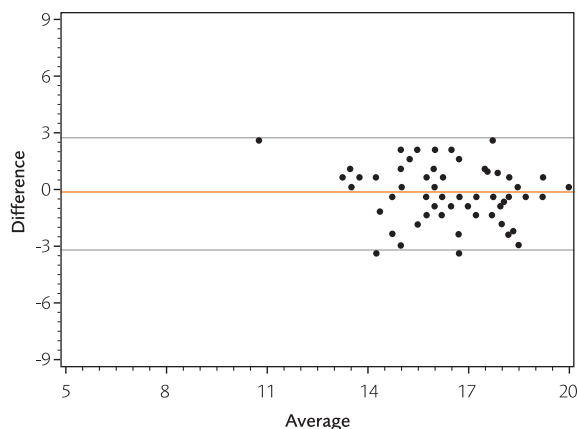
(a) Physical health domain



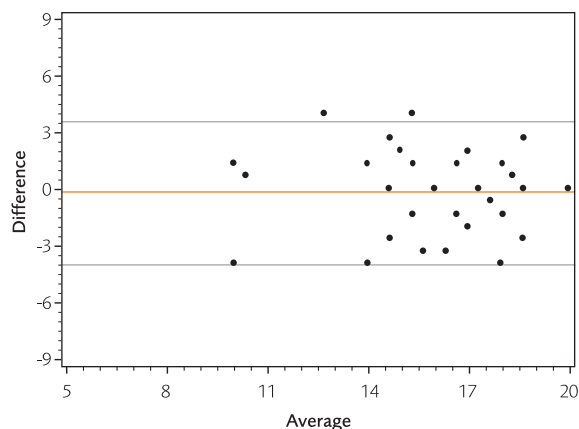
(b) Psychological health domain



(c) Social relationships domain



(d) Environment domain



Bland–Altman: The average of the two methods of administration is plotted against the mean difference between the two scores. Mid-line is the mean difference between the two methods of administration; upper and lower lines are two standard deviations from the mean difference.

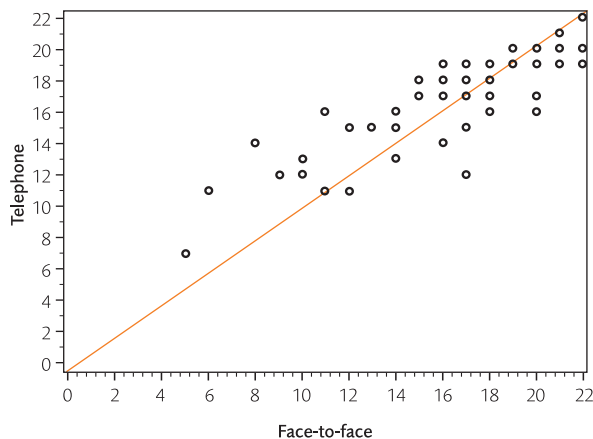
the kappa statistics are modest on individual items for the WHOQOL-BREF, the overall level of agreement is adequate because quality of life is variable on day-to-day enquiry and the majority of these interviews were completed more than eight weeks apart. The range of five responses on each item mean exact matches are less common. Each individual item contributes to a weighted score. It is the comparison of the domain scores that is more useful for establishing reliability. In addition, the kappas reported here are very similar to those reported in the validation manual when the two interviews

were exactly the same and conducted within two weeks of each other.⁴

The Pearson correlation coefficients for both instruments were high, which indicates a good strength of linear dependence between the two methods of administration (Figures 1 and 2). Furthermore the WHOQOL-BREF Bland–Altman graphs showed a good level of agreement and no indication of bias. The limits of agreement are relatively larger for the NEADL than the WHOQOL-BREF as a difference on score of four on the NEADL (of a total of 22) is more clinically

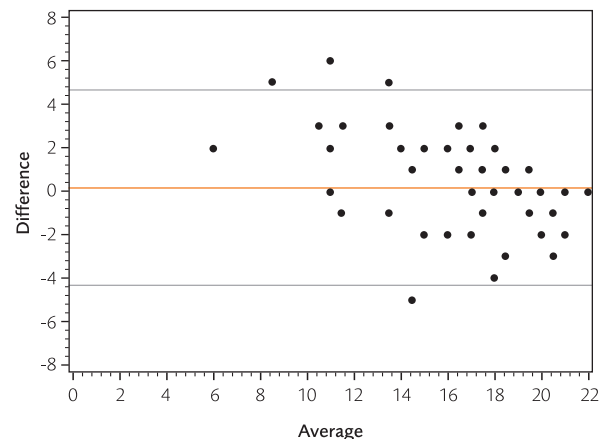
Figure 2. Graph for the total NEADL scores

(a) Correlation between face-to-face score vs telephone score



The score on the face-to-face method is plotted against the score from the telephone administered methods.

(b) Bland-Altman graph



The average of the two methods of administration is plotted against the mean difference between the two scores. Mid-line is the mean difference between the two methods of administration, upper and lower lines are two standard deviations from the mean difference.

important than a difference of three points on the WHOQOL-BREF. The Bland-Altman graph for the Total NEADL shows a tendency towards a greater negative difference the greater the average score becomes, which suggests a proportional error may occur as levels of disability rise. The very small mean differences in scores between the two modes of interviews and a high correlation coefficient, and the Bland-Altman graph of the Total NEADL indicate an acceptable level of agreement.

There are few studies that have compared telephone interviews with face-to-face interviews for research involving older people^{13,14} however many measures that have been compared produced promising results for a telephone-administered interview method. Comparison of face-to-face and telephone administration of a cognitive status measure for use in Alzheimer's disease generated a very high Pearson correlation coefficient ($r=0.904$), and substantial agreement between the two administration methods in older people (mean age = 73.9, SD = 8.8).¹⁴ Three quality of life measures in heart failure patients were noted to show no major differences comparing telephone and face-to-face administration methods (mean age = 57, SD = 12) as all three surveys pro-

duced high correlation coefficients, similar mean scores and satisfactory agreement.¹⁵ These studies show that the telephone-administration method yields valid results over a variety of measures.

A limitation to this present study is that all the participants received the telephone interview before the face-to-face interview as part of the procedure from the BRIGHT Trial since only those identified as disabled would receive a face-to-face interview. Despite the second interview being conducted within three months from the first interview, the order of the interviews may have accounted for the slightly higher face-to-face interview mean scores compared with the telephone interview. This may be due to a practise effect. A way to improve this would be to randomly select the order of the interviews for each participant so that half have the telephone interview first and the other has the face-to-face interview first. The interviews were also conducted farther apart than was ideal, potentially allowing health status and quality of life to change over time. The items of the WHOQOL-BREF, individually, showed only modest kappa statistics. The range of responses and the topics of each question, i.e. 'How satisfied are you with your personal relationships?'

are more likely to vary on a day-to-day basis than questions about function. This means the overall score may be more meaningful than each individual item. Thus the level of agreement between the face-to-face and telephone administration may be underestimated and may actually be closer than observed here.

This study is also limited by small sample size and the selected population. This was the more disabled group of those over age 74 years and, as such, exhibited a range of scores on both scales meaning the ceiling effects issues for function assessment of community populations were avoided.

Telephone interviews are more convenient, time efficient and economical to conduct, especially when there is a large number of people to assess. Quality of life and functional measures are two important aspects of health especially for older people and, with the growing proportion of the older population, efficient and valid ways of assessing the health of these people is important for future research, health practitioners and policy makers. Other researchers have shown that participants prefer to have a face-to-face interview, especially if there are any sensitive topics to be addressed.¹⁶ Also, interviewers would have a better opportunity to develop a rapport with the participant and observe any non-verbal signs such as body language, which is difficult to accomplish over the telephone.

In conclusion, telephone-administrated WHOQOL-BREF and NEADL can generally provide a reliable method to assess the quality of life and physical function in older people.

ACKNOWLEDGEMENTS

We sincerely thank all the practice nurses and participating general practices for their involvement in the BRIGHT Trial. We would also like to thank The University of Auckland for providing the Summer Studentship Grant, Elizabeth Robinson for providing statistical advice and reviewing the manuscript and C. Raina Elley for reviewing the manuscript.

FUNDING

The BRIGHT Trial is funded by an HRC project grant as part of a programme of gerontological work (06-068C).

COMPETING INTERESTS

None declared.

References

1. Statistics New Zealand. New Zealand's 65+ population: A statistical volume (2007). Retrieved December 24, 2008. Available from: <http://www.stats.govt.nz/analytical-reports/new-zealand-65plus-population.htm>
2. Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res.* 2004;13(2): 299–310
3. Chachamovich E, Trentini C and Fleck MP. Assessment of the psychometric performance of the WHOQOL-BREF instrument in a sample of Brazilian older adults. *Int Psychogeriatr.* 2007;19(4):635–646.
4. World Health Group. WHOQOL-BREF Introduction, administration, scoring and generic version of the assessment programme on mental health. The World Health Organization; 1996
5. World Health Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med.* 1998;28(3):551–558.
6. Fleck MP, Chachamovich E and Trentini C. Development and validation of the Portuguese version of the WHOQOL-OLD module. *Revista de Saude Publica.* 2006;40(5):785–791.
7. Pettit T, Livingston G, Manela M, Kitchen G, Katona C and Bowling A. Validation and normative data of health status measures in older people: the Islington study. *Int J Geriatr Psych.* 2001;16(11):1061–1070.
8. Young J, Robinson M, Chell S, Sanderson D, Chaplin S, Burns E et al. A prospective baseline study of frail older people before the introduction of an intermediate care service. *Health Soc Care Comm.* 2005;13(4):307–312.
9. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures. *Med Care.* 1997 May;35(5):522–37.
10. McClane KS. Screening instruments for use in a complete geriatric assessment. *Clin Nurs Spec.* 2006;20(4):201–207.
11. Hanmer J, Hays RD and Fryback DG. Mode of administration is important in US national estimates of health-related quality of life. *Med Care.* 2007;45(12):1171–1179.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;i:307–310
13. McPhail S, Lane P, Russell T, Brauer SG, Urry S, Jasiewicz J et al. Telephone reliability of the Frenchay Activity Index and EQ-5D amongst older adults. *Health and Quality of Life Outcomes.* 2009;7:48doi:10.1186/1477-7525-7-48.
14. Dal Forno G, Chiovenda P, Bressi F, Ferreri F, Grossi E, Brandt J et al. Use of an Italian version of the Telephone Interview for Cognitive Status in Alzheimer's disease. *Int J Geriatr Psychiatr.* 2006;21:126–133.
15. Bennett S, Oldridge N, Eckert G, Embree J, Browning S, Hou N et al. Comparison of quality of life measures in heart failure. *Nurs Res.* 2003;52(4):207–216.
16. Evans M, Kessler D, Lewis G, Peters TJ and Sharp D. Assessing mental health in primary care research using standardized scales: can it be carried out over the telephone? *Psycholog Med.* 2004;34(1):157–162.