ABSTRACT

The objective of the study was to assess the internal consistency and test-retest reliability of selected scales from four national and provincial surveys used to study the health status of community-dwelling seniors. Items on physical impairment, psychological well-being, activities of daily living, chronic health problems, social support, and stressful life events were selected from these surveys and administered in a group of 1,054 seniors, aged 65 and over, living in Wellington County, Ontario. Each scale exhibited high internal consistency (Cronbach's alpha ≥ 0.70) except for the stressful life events and social support scales. The intra-class coefficients (ρ) for all scales were > 0.80. In addition, almost all single items measures used in this study also exhibited high test-retest reliability (kappa > 0.80) except for measures of perceived health status (kappa = 0.69) and availability of help in crisis situation (kappa = 0.48).

ABRÉGÉ

L'étude avait pour objectif d'évaluer la cohérence interne et la fiabilité test-retest d'échelles sélectionnées dans quatre enquêtes nationales et provinciales visant à étudier l'état de santé de personnes âgées vivant en communauté. Des éléments relatifs aux handicaps physiques, au bien-être psychologique, aux activités quotidiennes, aux problèmes de santé chroniques, au soutien social, et aux événements stressants ont été sélectionnés dans ces enquêtes et administrés à 1 054 personnes âgées de 65 ans et plus, vivant dans le comté de Wellington en Ontario. Chaque échelle a montré une cohérence interne élevée (alpha de Cronbach ≥ 0,70) sauf dans le cas des échelles relatives aux événements stressants et au soutien social. Les coefficients intra-classe (ρ) pour toutes les échelles étaient supérieurs à 0,80. En outre, pratiquement toutes les mesures relatives à des éléments uniques utilisées dans cette étude affichaient une fiabilité test-retest élevée (kappa > 0,80) à l'exception des mesures de la perception de l'état de santé (kappa = 0,69) et de la disponibilité d'aide en cas de crise (kappa = 0,48).

How Reliable are Selected Scales from Population-based Health Surveys? An Analysis Among Seniors

Parminder Raina, PhD,[1] Brenda Bonnett, PhD,[1] David Waltner-Toews, PhD,[1] Christel Woodward, PhD,[2] Tom Abernathy, PhD[2,3]

Over the past decade, the number and type of studies assessing the health status and well-being of the elderly population have significantly increased.[1] National and provincial surveys, in particular, have produced rich sources of information about the health of seniors, as well as the general population. Such surveys often include a wide range of generic health status scales and the results, being in the public domain, serve as the basis for many subsequent public health and epidemiological studies. Although many of the scales in population surveys have undergone extensive testing, published literature contain very limited information about their internal consistency and test-retest reliability. Understanding the measurement properties of health scales may help to decide which scales are useful tools for reproducing the same or similar responses over time, and ensure that sources of variance due to content sampling are minimized.[2] In addition, such information can improve the reliability of questionnaires and increase the comparability of studies that use similar questions but in different population groups.

1. Department of Population Medicine, University of Guelph, Guelph, Ontario
2. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario
3. Central West Health Planning Information Network, Hamilton, Ontario

Correspondence and reprint requests: Dr. Parminder Raina, Assistant Professor, Department of Health Care and Epidemiology, Faculty of Medicine, University of British Columbia, L408, 4480 Oak Street, Vancouver, BC, V6H 3V4, Tel: 604-875-3572, Fax: 604-875-3569, E-mail: praina@sunnyhill.bc.ca

Research was conducted in the Department of Population Medicine at the University of Guelph, Guelph, Ontario, Canada N1G 2W5 and was supported by Human-Animal Studies Group, Waltham Centre, WALTHAM-ON-THE-WOLDS, Melton Mowbray, Leicestershire, England U.K.

This study focuses on a series of established health status scales that have been used to assess the social, physical and psychological health of the general population. The health status scales were selected from four recent population-based health surveys: the National Population Health Survey (NPHS),[3] the General Social Survey (GSS),[4] Ageing and Independence (AI),[5] and the Ontario Health Survey (OHS).[6] The objective of this study was to examine the internal consistency and test-retest reliability of single and multi-item health scales used in these surveys in an elderly population.

METHODS

Questionnaire pre-testing

Tables I and II show the items and response options of the scales used in the study. All the scales except "Activities of Daily Living" (ADL) and "Change in General Health from One Year Ago" appeared in four Canadian health surveys: NPHS, GSS, AI, and OHS. The ADL and Change in General Health from One Year Ago scales were adopted from the Medical Outcomes Study (MOS) Physical Functioning Measure.[7] Although the items in the four Canadian surveys were pretested extensively, the questions included in this study were also pre-tested to ensure that questions would not be misinterpreted and hence to prevent undesirable variations in data quality, response rates, and response validity.[1] The pretest resulted in no changes to the existing questions.

Data collection

A sample of 1,500 seniors stratified by age and sex were identified using the Ontario Ministry of Health's Registered Person Data

TABLE I
Multi-Item Health Status Scales and Responses Options

Scales*	# of items	Items Asked	Possible Ordinal/Categorical Responses
Physical Impairment: [OHS, GSS]	9	How is your eyesight (with glasses/contacts if needed)? How is your hearing (with hearing aid if needed)? Are you able to walk around without difficulty and without mechanical support? Are you able to walk around at all? Do you need mechanical support to be able to walk around? Do you need the help of another person to be able to walk? Do you need a wheelchair? How often do you use a wheelchair? Do you need the help of another person to get around in the wheelchair?	1 (excellent) - 5 (unable to see) 1 (excellent) - 5 (poor) Yes, No
Activities of Daily Living [MOS]	13	Are you able to do vigorous activities such as lifting heavy objects? Moderate activities such as moving a table, pushing a vacuum cleaner, scrubbing floors? Lift or carry groceries? Climb one flight of stairs? Climb several flights of stairs? Bend, kneel or stoop? Walk one block? Walk several blocks? Walk more than a mile? Get in and out of bed? Prepare your own meal? Take your own medicine? Bathe and dress yourself?	1 (can't do at all) - 4 (not at all limited)
Stressful Life Events [AI]	7	In the past 12 months have you changed or lost a job? Changed residences? Had a person move into or leave your home? Had a death in the family? Had a death in a close friend? Had a serious illness or injury? Had a family member or friend seriously ill or injured?	Yes, No
Social Supports (Family and Non-Family) [NPHS, GSS]	7	In the past 12 months, did you have contact with your daughters/step-daughters? Sons/step-sons? Grandchildren? Brothers/sisters? Other relatives? Close friends? Neighbours? For each, how often have you contacted them?	Yes, No 1 (everyday) - 5 (at least once a year)
Psychological Well-being [NPHS, OHS, GSS]	10	How would you describe yourself as being usually?	1 (happy and interested in life) - 5 (so unhappy that life is not worthwhile)
		How would you describe your usual ability to remember things?	1 (able to remember most things) - 4 (unable to remember anything at all)
		How would you describe your usual ability to think and solve problems?	1 (able to think clearly/solve problems) - 5 (unable to think/solve problems)
		How would you rate your feelings about your health? Job, or major activity? Finances? Housing? Family relations? Friendships?	1 (very satisfied) - 4 (very dissatisfied)
		How do you feel about your life as a whole at the present time?	1 (very satisfied) - 4 (very dissatisfied)
Chronic Conditions [NPHS, OHS, GSS]	20	Has a doctor ever told you that you have: e.g., allergies, arthritis or rheumatism, asthma, back problems, cerebral palsy, chronic bronchitis, etc.	Yes, No

* OHS=Ontario Health Survey; NPHS=National Population Health Survey; GSS=General Social Survey; AI=Aging and Independence; MOS=Medical Outcomes Study Physical Functioning Measure

Base (RPDB). Further description of the study is available elsewhere.[8] Briefly, from the sample of 1,500 seniors, 1,296 consented to participate and eventually 1,054 actually took part in the study (response rate=70%). The data were collected by telephone by trained interviewers and took approximately 30 minutes to complete.

Internal consistency and principal components analysis

For each scale used in the study, a mean combined score was generated by adding the numeric responses of each item. A missing score was assigned if more than half of the items on a scale were missing. If one half or fewer items were missing, a person-specific estimate (mean of the non-missing items) was used as the score.[9]

The internal consistency of the items within a multi-item scale was calculated as Cronbach's alpha (α).[10] An α value of 0.70 or greater was considered acceptable for each scale. Data from all 1,054 participants were used to examine the internal consistency of the scales. The principal components analysis was used as an additional tool to assess the unidimensionality of each multi-item scale in the questionnaire.[2]

Test-retest reliability

A convenience sample of 52 individuals from the 1,054 seniors was selected to measure the test-retest reliability of both single- and multi-item scales. The estimation of sample size necessary for measuring the test-retest reliability was based on: 1) the assumption of a test-retest reliability coefficient of 0.75, and 2) setting the lowest acceptable reliability coefficient at 0.50.

The retest telephone interview was conducted 15 days after the first interview and was identical to the original test questionnaire. The participants were interviewed by two different interviewers. A response was treated as missing in two situations: if a subject responded to a question at one time but their response to the same question was missing from the other, and if a response was missing from both the test and retest questionnaires.

TABLE II
Single-Item Health Status Scales and Response Options

Scales*	Items Asked	Possible Ordinal/Categorical Responses
Overall Quality of Life [NPHS]	How do you rate your quality of life?	1 (excellent) - 5 (poor)
Effect of Pain and Discomfort [NPHS]	Are you usually free of pain and discomfort? If no, which best describes the effect of the pain and discomfort you usually experience?	Yes, No 1 (does not prevent any activities) - 4 (prevents most activities)
Perceived Physical Health Status [NPHS]	Would you say your health is…	1 (excellent) - 5 (poor)
Level of Happiness [GSS]	At the present time, would you describe yourself as:	1 (very happy) - 4 (very unhappy)
Change in General Health from One Year Ago [MOS]	Compared to a year ago, how would you rate your health in general now? Are you:	1 (much better now than one year ago) - 5 (much worse now than one year ago)
Satisfaction with Life [GSS]	How do you feel about your life as a whole at the present time? Are you:	1 (very satisfied) - 4 (very dissatisfied)
Member of Voluntary Organization or Other Group [NPHS, OHS]	Are you a member of any voluntary organizations or associations such as church and school groups, labour unions, or social and civic clubs?	Yes, No
Active in Affairs of Group [NPHS, OHS]	How active are you in the affairs of the groups, clubs or organizations you belong to? If you belong to many, just think of the one in which you are most active. Are you:	1 (very active/attend most meetings) - 3 (not active/belong, but hardly ever go)
Availability of Confidant [NPHS, OHS]	Do you have someone you can confide in, talk to about yourself or your private feelings or concerns? Would you say…	1 (always) - 5 (never)
Availability of Help in Crisis Situation [NPHS, OHS]	Do you have someone you can really count on to help you out in a crisis situation, even though he/she had to go out of his/her way to do so? Would you say…	1 (always) - 5 (never)

* OHS=Ontario Health Survey; NPHS=National Population Health Survey; GSS=General Social Survey; AI=Aging and Independence; MOS=Medical Outcomes Study Physical Functioning Measure

The intra-class correlation coefficient (ρ) was calculated to account for a random interviewer effect since the two interviewers in the reliability study represented several interviewers in the population.[11] The estimate of ρ included variation between interviewers as a source of variance. A lower 100(1-α) confidence limit for each ρ was also calculated.[11]

For continuous single-item measures, test-retest reliability was also calculated using ρ. The test-retest reliability of each single item with binary and multi-level nominal responses was calculated using Cohen's kappa (κ), a measure of agreement over and above chance agreement.[12] The agreement between items with ordinal responses was calculated using weighted kappa (κ_w) with absolute weights.[13] The κ_w gives "partial credit" for close but not exact agreement.[14] We also calculated the 95% confidence interval associated with each k or κ_w.[13] Values of κ_w and κ between 0.00 and 0.20 are considered poor; those between 0.21 and 0.40, fair; those between 0.41 and 0.60, moderate; those between 0.61 and 0.80, substantial; and those between 0.81 and 1.00, perfect.[15]

RESULTS

Descriptive analysis of study sample

The sample was comprised of 51% females and had a mean age of 73 years (SD=6.3). The age distribution of the sample was 34% for 65-69 year olds, 32% for 70-74 year olds, 18% for 75-79 year olds and 16% for those 80 years and older. Twenty-eight percent of participants also had some form of post-secondary education. Sixty-four percent were married, 28% widowed and 8% were never married, separated or divorced. In addition, approximately 54% of the sample reported their gross household income to be below $25,000 per year. About 88% of participants reported being current non-smokers.[9]

The mean, standard deviation, and actual and possible ranges of the combined scores in each scale are shown on Table III. For all the scales, except for the ADL and stressful life events scale, the low range of the combined scores represented more favourable health responses to the items. For example, the low range of scores for physical impairment indicated less physical impairment. However, in the ADL and stressful life scales, the high range of the combined scores indicated high functioning in daily activities and low levels of stress, respectively.

Internal consistency and principle component analysis

Table IV shows the results from the internal consistency analysis of the multi-item scales. Internal consistency coefficients of 0.70 or greater were found for the following scales: physical impairment, psychological well-being, number of chronic diseases and ADL scales. These scales also exhibited unidimensionality on the principle component

analysis (not shown). Two scales, the stressful life events and family social supports, however, exhibited poor internal consistency. The internal consistency coefficient of the stressful life events scale was 0.23, suggesting that the scale may contain more than one dimension. The principal components analysis confirmed that the stressful life events items did not form a unidimensional scale but constituted three separate factors together accounting for 33% of the variance. New subscores were formed from each of these three factors, but each of these subscales still had very low consistency coefficient (less than 0.25). Therefore, in subsequent analyses each item was used separately because of lack of homogeneity of the scale.

The social support scale had a very low internal consistency coefficent of 0.26. The principle component analysis of 12 items indicated that three items on family social support and four items on non-family social support constituted separate factors which together accounted for 45% of the variance. Therefore, new scores for the two social support scales were calculated. The internal consistency of the family support scale and non-family social support scales increased to 0.61 and 0.48, respectively. The family support scale was marginally close to the cutoff of 0.70. The remaining five items related to social support (marital status, living with someone, availability of confidant, availability of support in crisis, and participation in voluntary organization) loaded on five different factors. Therefore, these items were not aggregated and were used as separate items in the subsequent analysis.

Test-retest reliability

The sample of 52 subjects was 61% male and had a mean age of 74 years (SD=6.6). In addition, the age distribution of the sample was 29% for 65-69 year olds, 35% for 70-74 year olds, 21% for 75-79 year olds and 15% for subjects over 80 years old.

The ρ along with lower 95% confidence interval of each scale, shown in Table IV, were found to be uniformly high (ρ>0.85). The reliability coefficients (κ or κ_w) for categorical items ranged from 0.48 (availability of help in crisis situation) to 0.88 (member of voluntary organization or other group) (Table V). The test-retest reliability (κ) coefficients for seven life event items ranged between 0.60 and 1.0 (not shown).

TABLE III
Descriptive Statistics and Combined Scores of Multi-Item Health Status Scales

Scales	# of items	Combined Score (n=1,054) Mean	SD	Actual Range	Possible Range*
Physical Impairment	9	5.85	2.67	3-21	3-25
Activities of Daily Living (ADL)	13	47.29	5.72	13-52	13-52
Stressful Life Events	7	12.95	0.97	9-14	7-14
Family Social Support	3	6.12	2.09	1-14	1-18
Non-Family Social Support	4	8.69	3.12	1-19	1-24
Psychological Well-Being	10	13.55	3.25	10-30	10-42
Chronic Conditions	20	2.96	1.92	0-12	0-20

* For each scale, except for the ADL and Stressful Life Events scale, low combined scores along the range represent more favourable health responses and high combined scores represent less favourable health responses. For example, a score of three for the Physical Impairment scale indicates less physical impairment and a score of 25 indicates more physical impairment. This was reversed for the ADL and Stressful Life Events scales. High scores for both scales represent high functioning in daily activities and low levels of stress, respectively.

TABLE IV
Internal Consistency (α) and Test-Retest Reliability (ρ) Coefficients of Multi-Item Health Status Scales

Scales	# of items	Internal consistency Cronbach's (α) (n=1,054)	Test-retest Reliability Intraclass correlation Coefficient (ρ) (95% Confidence Interval) (n=52)
Physical Impairment	9	0.79	0.95 (0.92-0.98)
Activities of Daily Living	13	0.92	0.93 (0.88-0.98)
Stressful Life Events	7	0.23	0.88 (0.81-0.95)
Family Social Support	3	0.61	0.89 (0.82-0.96)
Non-Family Social Support	4	0.48	0.94 (0.89-0.99)
Psychological Well-Being	10	0.70	0.93 (0.88-0.98)
Chronic Diseases	20	0.84	0.96 (0.93-0.99)

TABLE V
Test-Retest Reliability of Single-Item Scales

Scales	Test-retest reliability Kappa(κ)* or Weighted Kappa(κ_w)† and (95% Confidence interval) (n=52)
Overall quality of life	κ_w=0.86 (0.75-0.97)
Effect of pain and discomfort	κ_w=0.83 (0.73-0.93)
Perceived physical health status	κ_w=0.69 (0.54-0.84)
Level of happiness	κ_w=0.84 (0.75-1.00)
Compared to one year ago, rate your change of health	κ_w=0.84 (0.67-1.00)
Satisfaction with life	κ_w=0.83 (0.69-0.97)
Member of voluntary organization or other group	κ =0.88 (0.75-1.00)
Active in affairs of group	κ_w=0.87 (0.76-0.98)
Availability of confidant	κ_w=0.88 (0.74-1.00)
Availability of help in crisis situation	κ_w=0.48 (0.20-0.76)

* κ for nominal variables
† κ_w for ordinal variables

DISCUSSION

The items selected from the Canadian population health surveys represent generic health measures rather than disease-specific measures for assessing the health status of community-living seniors. The items were also intended to capture possible changes in health that would most commonly be encountered in a community-dwelling seniors population. Therefore, the key to understanding any subsequent findings from these surveys lies in the reproducibili-

ty and internal consistency of the measures being investigated.

In this study, Cronbach's α and principal components analysis were used to assess the unidimensionality and internal structure of each measure. A reasonably acceptable level of internal consistency ($\alpha \geq 0.70$) was found for all scales, except for the life events and overall social support scales. The low alpha level for these two scales appeared to be due to a lack of unidimensionality. Krause suggested that it may not be desirable to rely on a summated measure of several items of stressful life events[16] such as the one developed in our study. Instead, as shown in Krause's study, it may be preferable to examine the impact of specific life event items separately, such as death or illness in the family. Similarly, our study showed that measures which combine several dimensions of social support have lower internal consistency precisely because they fail to differentiate between different types of support, such as family versus non-family support. Thus, instead of an overall social support scale, two new scales were suggested, the family social support and non-family social support.

As described earlier, data from all 1,054 participants were used to examine the internal consistency of each multi-item scale. Due to the moderate response rate of 70%, the differences between respondents and non-respondents were assessed. A detailed analysis of respondents and non-respondents has been described in detail elsewhere.[8,17] Briefly, the sample included in this study underrepresented the females, old age groups and low income groups.[8,17] However, respondents were similar to non-respondents in many other characteristics such as marital status, education, physical activity and perceived health status.[8,17]

Our study found high test-retest reliability on all the scale measurements. High reliability estimates may have occurred as a result of a short time interval where subjects may remember their first response and repeat it, rather than answering the question *de novo*.[2] A learning bias may be particularly evident in questions that rely on opinion or memory,[2] such as perceived health status or availability of help in crisis. However, our study employed test-retest intervals of 15 days which has been proposed as the optimal time to minimize possible learning biases.[18,19]

The results of this study provide evidence that the use of health status questions based on established scales and items from national and provincial population surveys results in a reasonably good internal consistency and test-retest reliability when used in a population of seniors. However, this study used a convenience sample to measure the test-retest reliability. Future studies, therefore, will need to employ a more representative sample of seniors in order to establish how these health status measures perform on a general population of seniors.

REFERENCES

1. McHorney CA. Measuring and monitoring general health status in elderly persons: Practical and methodological issues in using the SF-36 Health Survey. *Gerontologist* 1996;36(5):571-83.
2. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* New York: Oxford University Press, 1989;104-27.
3. Statistics Canada, Canadian Centre for Health Information. National Population Health Survey. Ottawa: Ministry of Health Canada, 1992.
4. Statistics Canada, Housing, Family and Social Statistics Division. General Social Survey. Ottawa: Minister of Supply and Services Canada, 1991.
5. Statistics Canada, Ministry of National Health and Welfare. Ageing and Independence. Ottawa: Minister of Supply and Services, 1991.
6. Statistics Canada, Information, Planning and Evaluation Branch. Ontario Health Survey. Ottawa: Ministry of Health, 1989.
7. Stewart AL, Kamberg CJ. Physical functioning measures. In: Stewart AL, Ware JE Jr. (Eds.), *Measuring Functioning and Well-being: The Medical Outcomes Study Approach.* Durham, NC: Duke University Press, 1992;86-101.
8. Raina P, Waltner-Toews D, Bonnett B, et al. Influence of companion animals on the physical and psychological health of seniors: An analysis of a one-year longitudinal study. *J Am Geriat Soc* 1999 (in press).
9. Patrick DL, Bergner M. Measurement of health status in the 1990s. *Annu Rev Public Health* 1990;11:165-83.
10. Cronbach LJ. Co-efficient alpha in the internal structure of tests. *Psychometrika* 1951;16:297-334.
11. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York: John Wiley and Sons, 1986;1-32.
12. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurements* 1960;20:37-46.
13. Fleiss JL. *Statistical Methods for Rates and Proportions* 2nd ed. New York: John Wiley and Sons, 1981;188-236.
14. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurements* 1973;33:613-19.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
16. Krause N. Stressful life events and physician utilization. *J Gerontol* 1988;43(2):S53-S61.
17. Raina P. The Impact of Pet Ownership and Attachment on the Health and Health Care Use of the Elderly in Wellington County, Ontario [dissertation]. Guelph, Ontario: University of Guelph, 1995.
18. Boult L, Boult C, Pirie P, Pacala J. Test-retest reliability of a questionnaire that identifies elders at risk for hospital admission. *J Am Geriatr Soc* 1994;42:707-11.
19. Rubenstein L. Geriatric assessment: An overview of its impacts. *Clin Geriatr Med* 1987;3:1-15.

Received: May 12, 1998
Accepted: October 16, 1998