Efforts to establish the reliability of the Resident Assessment Instrument

Antonio Sgadari, John N. Morris[1], Brant E. Fries[2], Gunnar Ljunggren[3], Palmi V. Jónsson[4], Jean-Noël DuPaquier[5], Marianne Schroll[6]

Istituto di Medicina Interna e Geriatria, Università Cattolica del Sacro Cuore, Largo A. Gemelli, 8, 00168 Rome, Italy
[1]Hebrew Rehabilitation Center for the Aged, Division of Social Gerontology, Boston, USA
[2]RAI Unit, Sabbatsberg Sjukhus, Stockholm, Sweden
[3]Institute of Gerontology and School of Public Health, University of Michigan and Veterans Administration Medical Center, Ann Arbor, MI, USA
[4]Me-Ti SA, Carouge, Switzerland
[5]Department of Geriatrics, Reykjvik City Hospital, Reykjavik, Iceland
[6]Department of Geriatrics, Kommunehospitalet, Copenhagen, Denmark

Address correspondence to A. Sgadari. Fax: (+39) 6 305 1911. E-mail: A.SGADARI@caspur.it

Abstract

Background: since its original implementation in the USA, the Resident Assessment Instrument (RAI) has been used in many countries in languages other than English. This paper describes the efforts that have been made to test the inter-rater reliability of the core set of items forming the minimum data set items in the USA and in non-English speaking countries (Denmark, Iceland, Italy, Japan, Sweden and Switzerland).
Results: a large proportion (from 70 to 96%) of the items in the RAI achieved an adequate to excellent level of reliability, with no substantial differences across countries. The RAI met the standard for good reliability (i.e. a κ value of 0.6 or higher) in crucial areas of functional status, such as memory, activities of daily living self-performance and support, and bowel and bladder continence in most of the countries. Indicators of mood and behavioural problems achieved adequate reliability levels of 0.4 or higher.

Keywords: assessment instruments, minimum data set, nursing home, reliability

Introduction

There is general agreement that developing reliable data on the functional and mental status of nursing home residents is a task that challenges traditional approaches to measurement. A large proportion of nursing home residents have some level of cognitive impairment and/or communication difficulties. Furthermore, their abilities and their status vary during the day and across the week [1].

The RAI was designed to produce reliable data on even the most abstract concepts, using mutually exclusive response categories, standardized definitions, directions for inquiring about resident status over a relevant period of time and multiple sources of information [2]. In its development, three different field tests were conducted which, after dropping or revising more than half of the original items, led to the final version of the RAI in 1991 [3, 4].

The success of the RAI in the USA made it attractive to the international community of investigators in geriatrics and gerontology and it started being adopted in European countries and in Japan. National versions were translated into several languages: Czech, Danish, Dutch, Finnish, French, German, Italian, Icelandic, Japanese, Norwegian, Spanish and Swedish. Soon it became clear that RAI could represent a sort of 'Esperanto'—a common language that could enable researchers from different countries and cultures to share data, compare populations and carry out cross-national studies on elderly populations.

The success of this international comparative research relies heavily on the quality of the data that are collected. As a preliminary step, in most countries an independent 'back-translation' to English was accomplished to assure that the translation process had not caused major changes in the wording, phrasing and meaning of the original RAI items. Then, assessment staff were trained to perceive the need for accurate,

reliable data, to understand the roles of the different participants in the data collection process and to complete the RAI form accurately and efficiently. Eventually, in several countries studies were carried out to verify that the national version of the RAI led to uniform and reliable assessment.

The aim of the present paper is to describe and accumulate evidence from the efforts that have been made to test the inter-rater reliability of RAI items in Denmark, Iceland, Italy, Japan, Sweden, Switzerland and the USA.

Methods

The results presented in this paper are based on dual assessments completed by licensed nursing personnel. All nursing assessors were trained by project investigators. In each facility, two nurses were assigned a common group of residents to assess using the RAI. The nurses independently performed their assessments and were prohibited from discussing their residents or their findings before the conclusion of the study.

An overview of the procedures used in the reliability studies in seven countries are shown in Table 1. In European countries and Japan the residents were randomly selected from a roster. In the USA, since the final RAI reliability testing also included the validity testing of the resident assessment protocols, residents were stratified according to the presence or absence of a specific problem under review before being selected.

The largest number of assessments was performed in Japan ($n = 129$), the lowest in Iceland ($n = 24$). In Switzerland and Sweden the two nurses assessed the same resident within 2 days, while in other countries the time frame was, on average, 7 days. In Japan the assessments were carried out over 2 weeks.

Analytic approach

In each country, the reliability of the RAI items was assessed by comparing RAIs completed by two trained nurses who independently evaluated the same residents. The reliability testing differed somewhat since the interviewers assessed the resident at the same time and concurrently accessed other sources of information. Both assessors employed the recommended

information gathering procedures and the same RAI items, response categories and definitions.

Reliability estimates were based on the congruence between judgements of the two assessors using the weighted κ [5, 6] and the Spearman–Brown intraclass correlation coefficient [7]. These measures generally provide a more conservative estimate of reliability than either simple correlation or percent agreement because they adjust for chance agreement between the two assessors. It can be demonstrated that the intraclass correlation coefficient corresponds to κ when dichotomous variables are considered and to weighted κ when applied to categorical variables [8].

For the sake of consistency through the paper the term 'κ' will be used in the text. The κ statistic can assume values between 1 and -1. It is negative when agreement is poorer than chance and positive when agreement is better than chance. A κ of 1 denotes perfect agreement. According to Fleiss [5], κ values lower than 0.4 indicate poor reliability. When the κ statistic is in the range 0.40–0.75, reliability is considered adequate and a value above 0.75 is considered evidence of excellent reliability. In a limited number of instances, κ was not calculated because the item had little or no variance.

Results

The results demonstrate that RAI item reliabilities are adequate for research purposes. The percentages of all 300+ RAI items achieving a κ coefficient of 0.4 or higher are the following: 96% in Denmark, 95% in Iceland, 90% in Italy, 80% in Japan, 76.1% in Sweden, 97% in Switzerland and 88% in the USA.

In addition, a reasonable percentage of items attained excellent reliability (0.75 or higher) in most countries. This included 84% of the items in Switzerland and between 49 and 34% in Italy, Denmark, Iceland and the USA. The lowest percentages were found in Sweden and Japan (31.3 and 16.7%, respectively).

Sections of the RAI contain multiple items describing crucial domains of a resident's functioning. The reliability of items by these domains are reported in Table 2. For instance, the memory/memory recall indicator represents the average of the six individual items which investigate short- and long-term memory. The key indicators of functional status meet the

Table 1. Characteristics of reliability testing studies

	Denmark	Iceland	Italy	Japan	Sweden	Switzerland	USA
Sample selection	Random	Random	Random	Random	Random	Random	SR
No. of facilities	8	3	2	15	3	19	13
No. of assessments	74	24	82	129	46	87	123
Time frame (days apart)	9-12	7	7	14	1-2	Same day	7

SR, stratified random.

Table 2. Average reliability of crucial sets of indicator items

Indicators	Denmark	Iceland	Italy	Japan	Sweden	Switzerland	USA
Memory/memory recall	0.68	0.80	0.88	0.70	0.64	0.88	0.69
Cognitive skills	0.59	0.81	0.88	0.56	0.47	0.86	0.88
Hearing	0.76	0.39	0.83	0.53	0.65	0.80	0.84
Communication/understanding	0.83	0.63	0.82	0.49	0.67	0.89	0.66
Vision patterns	0.67	0.80	0.74	0.61	0.50	0.90	0.62
ADL self performance	0.68	0.74	0.90	0.61	0.75	0.87	0.92
ADL support provided	0.74	0.76	0.84	0.57	0.75	0.85	0.87
Bowel and bladder continence	0.74	0.58	0.95	0.67	0.74	0.93	0.91
Indicators of initiative	0.71	0.70	0.68	0.53	0.38	0.86	0.58
Unsettled relationships	0.79	0.84	0.63	0.46	0.54	0.90	0.49
Mood distress	$-^{a}$	0.69	0.60	0.45	0.63	0.93	0.44
Behaviour problems	0.72	0.59	0.59	0.34	0.60	0.81	0.63
Nutritional status	$-^{a}$	0.78	0.65	0.43	0.54	0.99	0.69

ADL, activities of daily living.
[a] κ coefficient could not be calculated: very low prevalence, no variance or data not available.

standard for excellent reliability in the vast majority of the countries, with Sweden and Japan constantly showing lower, but still adequate, figures. Nearly all the areas relevant to understanding a resident's status attain reliability coefficients of 0.6 or higher in all the countries.

A minimum κ value of 0.61 and of 0.58 was achieved in activities of daily living (ADL) self-performance and bladder/bowel continence, respectively—areas that deeply affect residents' functional status and their risk of entering and remaining in a nursing home. The indicators of mood distress, behaviour problems and nutritional status achieved lower reliability values in a consistent way across countries.

In Figure 1, the overall reliability of RAI items is represented. The distribution of κ values in the range −1 to 1 has a very similar pattern in Denmark, Iceland, Italy and USA, with only a small proportion of the items (4-12%) having poor reliability in comparison to a large percentage achieving adequate (41-55%) or excellent reliability (35-49%). Sweden and Japan had figures that were somewhat lower: the percentages of items with excellent κ values were only 21 and 17%, while those with poor reliabilities were 29 and 20%, respectively. Still, a very high proportion of items (50 and 63%) reached adequate reliability. Finally, a different picture was observed in Switzerland, where most items achieve inter-rater reliability coefficients of 0.75 or higher.

Discussion

The results show that the vast majority of RAI items achieve adequate to excellent reliability in all the countries. In many areas which are critical to describe resident's functional status, RAI items exhibit excellent reliabilities in most samples, including such items as ADL self performance, ADL support and hearing.
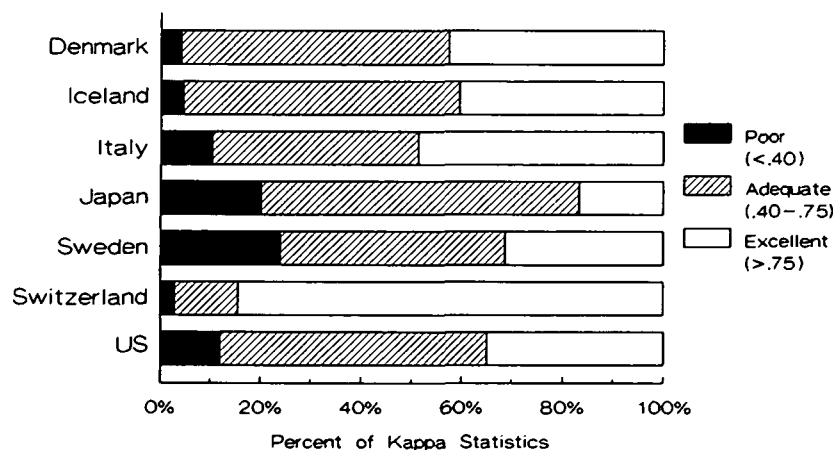
Figure 1. Overall reliability of minimum data set items.

Adequate reliabilities were attained in almost all countries for such items as memory/memory recall, cognitive skills, communication understanding and bowel/bladder incontinence. Differences in how the reliability studies were performed may account for the consistently higher figures observed in Switzerland. It is also noteworthy that, even for indicators that achieved poorer reliabilities, we observed a similar phenomenon across countries. This suggests that these less favourable results were intrinsic to the original items and not due to cultural differences nor were they introduced by the translation process. These findings demonstrate that data collected using the RAI instrument and according to the specified information-gathering protocols are useful both for clinical and research purposes.

The RAI includes more than 300 items addressing key elements of a nursing home resident's functioning, health conditions, services received, preferences and care needs. All these data elements are essential to the development of an individualized care plan. Also, to prove the utility of the RAI in the clinical setting, these data must be reliable. Our findings show that a large proportion of these items are reliable, with slight differences across countries.

The implications for scientific research are important as well. It has been argued that RAI protocols for information gathering lead to error in measurements [9]. On the contrary, our results confirm previous findings demonstrating that RAI can be used as a research tool. Data collected using RAI are as reliable as other research instruments, the reliability of which has been established in field trials [4].

Since its introduction in the USA, RAI has been used in many countries and national versions of the instrument are currently used in an increasing number of long-term care institutions all over the world. Thus, the potential exists for RAI to become a common language that allows researchers from different countries to understand one another. Comparing elderly residents living in long-term care institutions in different countries has historically been difficult since the same names of institutions in each country vary greatly and have different meanings in different cultures. Given this further source of heterogeneity, it is even more critical that data used for the purposes of cross-national comparisons are solid and reliable and based on the resident [10].

The major finding of the present paper is that, although some differences were found in the patterns of reliabilities, data collected using RAI are reliable and similarly so across numerous measurement domains in all the countries participating in the study. Therefore, once the potential case-mix differences in the composition of national samples are corrected for, RAI provides researchers with the unique opportunity of comparing the prevalence and incidence of a number of residents' outcomes as well as the facilitating comparisons of care practices.

Analytic studies investigating specific topics are expected in the near future and some are reported in this supplement, demonstrating that RAI and cross-national data bases are useful in scientific research.

References

1. Hawes C, Phillips C, Mor V et al. MDS data should be used for research. Gerontologist, 1992; 32: 563-4.

2. Morris JN, Hawes C, Murphy K et al. Resident Assessment Instrument Training Manual and Resource Guide. Natwick, MA: Eliot Press, 1991.

3. Morris JN, Hawes C, Fries BE et al. Designing the National Resident Assessment Instrument for nursing homes. Gerontologist 1990; 30: 293-302.

4. Hawes C, Morris JN, Phillips CD et al. Reliability estimates for the Minimum Data Set for nursing home residents assessments and care screening (MDS). Gerontologist 1995; 35: 172-8.

5. Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed. New York: John Wiley, 1981.

6. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968; 70: 213-20.

7. Fleiss JL. The Design and Analysis of Clinical Experiments. New York, John Wiley, 1986.

8. Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Ment Dis 1976; 163: 307-17.

9. Teresi JA, Holmes D. Should MDS data be used for research? Gerontologist 1992; 32: 148-9.

10. Fries BE, Schroll M, Hawes MC, Gilgen R, Jonsson PV. Approaching cross-national comparisons of nursing home residents. Age Ageing 1997; 26 (suppl. 2): 13-8.