Standardizing Assessment of Elderly People in Acute Care: The interRAI Acute Care Instrument

Len C. Gray, MD, PhD,[a] Roberto Bernabei, MD,[bc] Katherine Berg, PhD, PT,[d] Harriet Finne-Soveri, MD, PhD,[e] Brant E. Fries, PhD,[fg] John P. Hirdes, PhD,[hi] Pálmi V. Jónsson, MD,[jk] John N. Morris, PhD,[l] Knight Steel, MD,[mn] and Sergio Ariño-Blasco, MD[o]

OBJECTIVES: To examine the frequency distributions and interrater reliability of individual items of the interRAI Acute Care instrument.

DESIGN: Observational study of a representative sample of older inpatients; duplicate assessments conducted on a subsample by independent assessors to examine interrater reliability.

SETTING: Acute medical, acute geriatric and orthopedic units in 13 hospitals in nine countries.

PARTICIPANTS: Five hundred thirty-three patients aged 70 and older (mean age 82.4, range 70–102) with an anticipated stay of 48 hours or longer of whom 161 received duplicate assessments.

MEASUREMENTS: Sixty-two clinical items across 11 domains. Premorbid (3-day observation period before onset of the acute illness) and admission (the first 24 hours of hospital stay) assessments were conducted.

RESULTS: The frequency of deficits exceeded 30% for most items, ranging from 1% for physically abusive behavior to 86% for the need for support in activities of daily living after discharge. Common deficits were in cognitive skills for daily decision-making (38% premorbid, 54% at admission), personal hygiene (37%, 65%), and walking (39%, 71%). Interrater reliability was substantial in the premorbid period (average $\kappa = 0.61$) and admission period (average $\kappa = 0.66$). Of the 69 items tested, less than moderate agreement ($\kappa < 0.4$) was recorded for six (9%), moderate agreement ($\kappa = 0.41–0.6$) for 14 (20%), substantial agreement ($\kappa = 0.61–0.8$) for 40 (58%), and almost perfect agreement ($\kappa > 0.8$) for nine (13%).

CONCLUSION: Initial assessment of the psychometric properties of the interRAI Acute Care instrument provided evidence that item selection and interrater reliability are appropriate for clinical application. Further studies are required to examine the validity of embedded scales, diagnostic algorithms, and clinical protocols. **J Am Geriatr Soc 56:536–541, 2008.**

Key words: elderly; comprehensive geriatric assessment; hospital; acute care; electronic medical records

From the [a]Academic Unit in Geriatric Medicine, School of Medicine, University of Queensland, Brisbane, Australia; [b]Department of Internal Medicine and Geriatrics, Università Cattolica del Sacro Cuore, Rome, Italy; [c]Department of Gerontology, Geriatrics, and Rehabilitative Medicine, A. Gemelli University Hospital, Rome, Italy; [d]Department of Physical Therapy, University of Toronto, Toronto, Ontario, Canada; [e]National Research and Development Centre for Welfare and Health, Helsinki, Finland; [f]Institute of Gerontology, University of Michigan, Ann Arbor, Michigan; [g]Health Systems Research, Veterans Affairs Healthcare System, Ann Arbor, Michigan; [h]Homewood Research Institute, Guelph, Ontario, Canada; [i]Department of Health Studies and Gerontology, University of Waterloo, Waterloo, Ontario, Canada; [j]Department of Geriatrics, Landspitali University Hospital, Reykjavik, Iceland; [k]Department of Geriatrics, School of Medicine, University of Iceland, Reykjavik, Iceland; [l]Institute for Aging Research, Hebrew Senior Life, Boston, Massachusetts; [m]Department of Geriatric Medicine, Hackensack University Medical Center, Hackensack, New Jersey; [n]Institute for Incentives in Health Care, Guttenberg, New Jersey; [o]Department of Geriatrics, Hospital General of Granollers, Barcelona, Spain.

Address correspondence to Professor Len Gray, Director, Academic Unit in Geriatric Medicine, School of Medicine, The University of Queensland, Princess Alexandra Hospital, Ipswich Road, Brisbane, Qld 4102, Australia. E-mail: len.gray@uq.edu.au

Standardized clinical data systems have the potential to improve the effectiveness, efficiency, and administration of health systems by facilitating electronic records; creating a common language among clinicians; and providing data for outcome assessment, case mix analysis, quality monitoring, benchmarking, and planning.

Frail older patients typically have multiple complex problems that are chronic in nature and often require the ongoing support of several services. In that context, electronic medical record systems incorporating standardized assessment data can assist in the coordination of care.

Standardized approaches to assessment are in frequent use. Typically, an assessment schedule comprises a compilation of individual instruments that focus on each relevant clinical domain (e.g., the Mini-Mental State Examination for dementia screening, the Barthel Index for activities of daily living (ADLs)). This approach has weaknesses related

to lack of proven utility of instruments across care settings, the potential for repetition of data items in slightly different formats in each instrument, the differing approaches to data collection (e.g., interview, observation), and gaps that occur in relation to some clinical domains for which appropriate instruments are not available.

Several assessment schedules adopt a more-integrated approach, but these instruments are designed primarily for community-dwelling individuals.[1,2] None of these approaches have been used to specifically construct assessment instrumentation that transcends care settings.

With these concepts in mind, in 2005, the interRAI international collaborative released a fully integrated suite of assessment instruments with versions for mental health, postacute, palliative, home, and long-term care.[3] These instruments share common data items, assessment methods, outcome scales, and decision support tools for care planning. An acute care version—the subject of this study—was released in 2006.

STUDY AIMS

The study reported here was designed to examine the frequency distributions and to establish the interrater reliability of the individual assessment items within the interRAI Acute Care instrument (interRAI AC). Frequency distributions were examined particularly to ensure that the prevalence of items in the assessment was sufficiently great to justify their inclusion.

METHOD

Participating Organizations

Members of the interRAI consortium approached hospitals within their country of origin to participate in the trial. To be eligible, the organization was required to recruit at least 50 acute care cases (of which 15 were to be duplicated by a second assessor) in an acute medical, acute geriatric, or orthopedic unit.

Subjects

Specific eligibility criteria were as follows: aged 70 and older; admitted to an acute general medical, acute geriatric, or orthopedic unit; and anticipated to remain in the hospital 48 hours or longer. Patients admitted to intensive or coronary care units and unconscious patients were excluded.

Instrumentation

The interRAI AC represents a major revision of an earlier instrument.[4] The current version was designed to achieve compatibility with the interRAI suite, creating the possibility of data transfer at admission from, and discharge to, other services using other instruments in the suite. With information collected during premorbid, admission, review, and discharge observation periods, the new version supports entire inpatient episodes, as well as single geriatric consultation services.

The instrument comprises a set of recorded observations, a detailed instruction manual, and a set of algorithms used to interpret the observations. A panel of 15 expert clinicians and measurement scientists (including 8 geriatri-

cians) who met on numerous occasions in person and by teleconference selected the items.

There are 62 clinical items across 11 domains, including cognition, communication, mood and behavior, ADLs, instrumental activities of daily living (IADLs), continence, nutrition, falls, medical diagnoses, medications, advance directives, and discharge potential.

The admission assessment is completed 24 hours after admission, at which time information is also collected describing the person's status in the premorbid period—defined as the 3 days before the onset of the acute illness. This premorbid assessment provides a profile of the baseline status of the individual and provides important referencing information to appraise the potential for improvement after recovery or rehabilitation.

The instrument has been designed to incorporate diagnostic and risk screens for adverse events that might occur during an inpatient stay or in the postacute period. For example, there are diagnostic screens for dementia, delirium, and malnutrition and risk profiling for falls, pressure ulcers, and institutional care.

An array of scales that serve as severity and outcome measures, and in some cases as diagnostic screens, can also be calculated from the instrument, including ADLs, IADLs, cognition, communication, pain, and body mass index. These are based on scales developed by interRAI for other care settings.[5–7]

Finally, algorithms designed to identify patients most likely to benefit from preventive or curative interventions have been devised based on research in acute care and other clinical settings. Together with guidelines that suggest approaches to further assessment and management strategies, these are entitled Clinical Assessment Protocols (CAPs) and are an important feature of all interRAI instruments.[8]

Figure 1 illustrates the schema for data collection and the array of clinical and administrative outputs across an acute care episode.

Copies of the data set and further descriptions of the instrument can be obtained from interRAI through its Web site.[3]

Assessors

At each site, two assessors were trained in the administration of the interRAI AC. After a minimum 1-day training session, two practice assessments were performed and the results reviewed with the assessor by the trainer. A comprehensive training manual was provided to each assessor.

Procedures

Depending on daily caseload, sequential or randomly selected cases were recruited. Although the full schedule includes review and discharge items, in this study, only the premorbid and assessment items were collected. The premorbid and admission periods were selected for two reasons. First, all of the review and discharge items are included in the admission item set. Second, the patients' status is more complex and less stable at admission, and this was considered to be the most challenging period in which to test interrater reliability of the items. Assessments were conducted between 24 and 48 hours after admission, and the relevant observation period was the first 24 hours of the
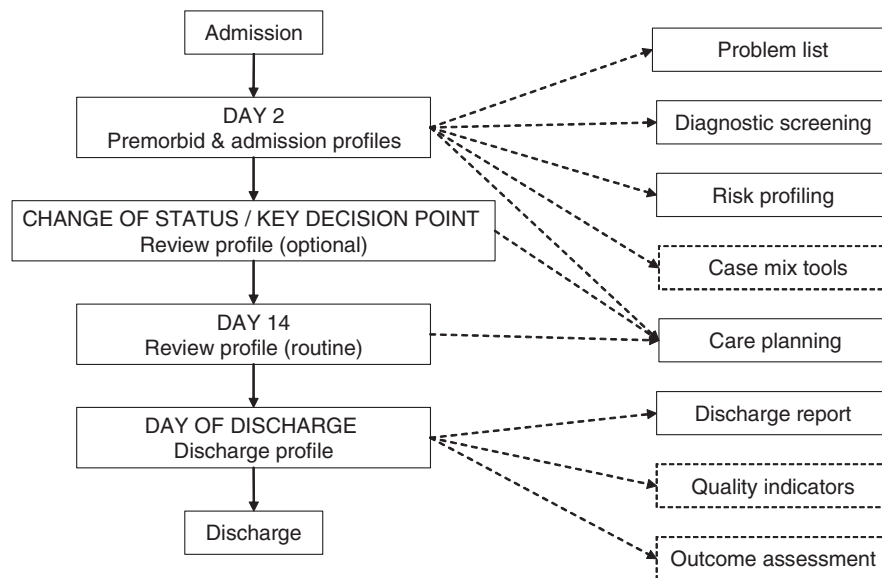
Figure 1. Scheme illustrating data collection periods and clinical and administrative outputs of the interRAI Acute Care instrument.

stay. At a fully developed site, 35 single assessment cases were conducted along with an additional 15 duplicate assessments (within 4 hours) for the interrater reliability study. In all duplicate assessment procedures, the assessors were blinded to the earlier assessment results.

The method of clinical observation used in all interRAI instruments was applied as follows. Health professionals collect data using a combination of subject interview, primary and secondary observation (interviewing nursing staff and relatives), and chart review. Assessors interview primary caregivers to obtain collateral information. Assessors are trained to distill information from a variety of sources and to be cautious about relying solely on patient report. This approach has been shown repeatedly to produce good reliability across interRAI instruments. Assessments required 20 to 90 minutes, depending on case complexity.

Analysis

Analysis was conducted using SPSS version 15.0 (SPSS Inc., Chicago, IL) at the University of Queensland and at the Hebrew Senior Life Center in Boston. Frequency distributions were examined by calculating the proportion of cases with abnormal findings and subsequently compared between each ward type (acute medical, acute geriatric, and orthopedic). Interrater reliability was evaluated by calculating weighted kappa coefficients using Fleiss-Cohen weights. Kappa values of 0.40 or higher indicate acceptable reliability.[9] The descriptors of strength of agreement provided by Landis and Koch are used in this paper (e.g., the κ range 0.61–0.80 is described as "substantial").[10]

RESULTS

Thirteen sites from Australia (2), Canada, Finland, Iceland, Italy, Japan (4), Norway, Mexico, and Spain participated. They recruited a total of 553 cases, including 146 duplicate cases for premorbid data and 161 cases for admission data to be used for examination of interrater reliability. Duplicate assessments were conducted at 10 of the 13 sites. There

were no significant differences in frequency distributions between the duplicate and other cases with regard to key demographic or clinical variables (such as age, sex, ward type, and cognitive skills for daily living). All assessors were registered nurses except at one site where a medical practitioner was involved.

Missing data for clinical items were rare, with fewer than 1% missing in any variable. There were 249 (45%) subjects in acute medical, 197 (36%) in acute geriatric, and 104 (19%) in acute orthopedic units. Two-thirds (66%) were female. The mean age ± standard deviation was 82.4 ± 7.2 (range 70–102); 37% were currently married; and 36% usually lived alone and 14% in a group setting (including 5% in a nursing home). Thirty percent had been in the hospital in the previous 90 days.

Frequency Distributions

Deficits (or clinical abnormalities) were common in the majority of clinical domains in both the premorbid and admission observation periods. Of the 30 premorbid observation variables, deficits were present in an average of 38% of subjects, ranging from 1.1% for the presence of physically abusive behavior to 78% for the item related to subnormal activity levels (<2 h/d) before the onset of the current acute illness. In five items, the proportion was less than 5%.

Similarly, deficits were present in 38% of the 39 admission items, ranging from 0.9% for the physical abuse item to 86% for the item related to the need for ADL and IADL support after discharge. In four items, the proportion was less than 5%.

Of the 18 items recorded in both the premorbid and admission periods, there was a general increase in the proportion of subjects with abnormal findings (range − 3 − 32% (Table 1 and Figure 2)). There was little observed change in behavioral symptoms, for which the incidence was low. Marked declines in ADL function were particularly evident. Deficits were common across all three clinical settings, although the incidence tended to be

Table 1. Frequencies and Levels of Agreement of Clinical Observations

Item Name	Premorbid*			Admission*		
	Abnormal, % (n = 553)	κ (n = 146)	95% CI	Abnormal, % (n = 553)	κ (n = 161)	95% CI
Cognitive skills for daily decision-making	37.7	0.89	0.84–0.93	53.8	0.83	0.72–0.90
Memory and recall ability	40.6	0.72	0.63–0.78	46.2	0.73	0.65–0.80
Easily distracted				29.2	0.76	0.64–0.85
Episodes of disorganized speech				23.4	0.67	0.52–0.79
Mental function varies over course of day				25.6	0.70	0.56–0.80
Acute change in mental function from baseline				23.4	0.65	0.50–0.79
Making self understood (expression)				34.1	0.85	0.72–0.92
Ability to understand others (comprehension)				36.1	0.86	0.72–0.92
Vision				44.2	0.63	0.47–0.81
Hearing				39.5	0.64	0.48–0.78
Little interest or pleasure in things you would normally do	30.7	0.61	0.45–0.75	37.4	0.66	0.49–0.78
Anxious, restless, or uneasy	32.5	0.57	0.46–0.78	38.7	0.59	0.45–0.70
Sad, depressed, or hopeless	29.9	0.66	0.48–0.77	35.0	0.56	0.36–0.67
Verbal abuse	1.6	0.65	0.55–0.74	1.3	0.32	0.19–0.44
Physical abuse	1.1	0.49	0.37–0.60	0.9	1.00	.
Socially inappropriate or disruptive behavior	2	0.79	0.72–0.85	1.8	0.44	0.31–0.55
Resists care	3.3	0.27	0.12–0.41	2.5	0.22	0.08–0.36
Personal hygiene	37.3	0.75	0.63–0.83	65.0	0.79	0.69–0.87
Walking	38.7	0.70	0.51–0.83	70.7	0.80	0.71–0.89
Toilet use	34.7	0.70	0.53–0.81	65.4	0.85	0.77–0.92
Eating	26.6	0.59	0.33–0.74	44.6	0.63	0.51–0.79
Bathing	54.6	0.59	0.44–0.73			
Bed mobility	23.9	0.71	0.55–0.87	49.9	0.83	0.74–0.89
Used a wheelchair	20.7	0.86	0.75–0.95	18.0	0.17	0.02–0.39
Activity level (<2 hours)	77.5	0.66	0.54–0.77			
Stamina	63.2	0.61	0.50–0.77			
Confined to bed				27.0	0.74	0.66–0.80
Meal preparation	66.9	0.70	0.58–0.79			
Ordinary housework	75.7	0.57	0.39–0.68			
Managing finances	60.5	0.63	0.52–0.74			
Managing medications	55.1	0.66	0.51–0.77			
Phone use	38.1	0.63	0.48–0.78			
Stairs	62.3	0.53	0.40–0.66			
Shopping	71.1	0.62	0.48–0.79			
Transportation	69.9	0.57	0.42–0.68			
Balance: seated to standing				61.1	0.72	0.59–0.82
Balance: turned around and faced opposite direction				62.3	0.68	0.56–0.78
Bladder continence	45.6	0.56	0.38–0.69	54.9	0.61	0.47–0.73
Urinary collection device	2.4	0.17	− 0.02–0.49	19.8	0.70	0. 53–0.83
Bowel continence	22.1	0.63	0.47–0.88	33.5	0.57	0.40–0.70
Falls				38.1	0.73	0.59–0.84
Pain frequency				48.0	0.74	0.62–0.83
Pain intensity				47.0	0.67	0.50–0.77
Pain consistency				47.0	0.68	0.54–0.77
Weight loss				18.3	0.62	0.50–0.71
Mode of nutritional intake				24.3	0.78	0.59–0.90
Pressure ulcer before admission	5.1	0.24	0.11–0.36			
Pressure ulcer (including skin redness)				8.9	0.71	0.37–0.89
Prior community service (using service)				51.0	0.68	0.53–0.78
Assistance needed with activities of daily living and instrumental activities of daily living				86.4	0.56	0.33–0.71
Medical social supports needed				74.1	0.41	0.26–0.59

*Premorbid refers to a 3-day period before the onset of the current acute illness. Admission refers to the initial 24 hours from the time of admission to hospital. Missing data <1%. Data collected 24 to 48 hours after admission to hospital, during daytime working hours.
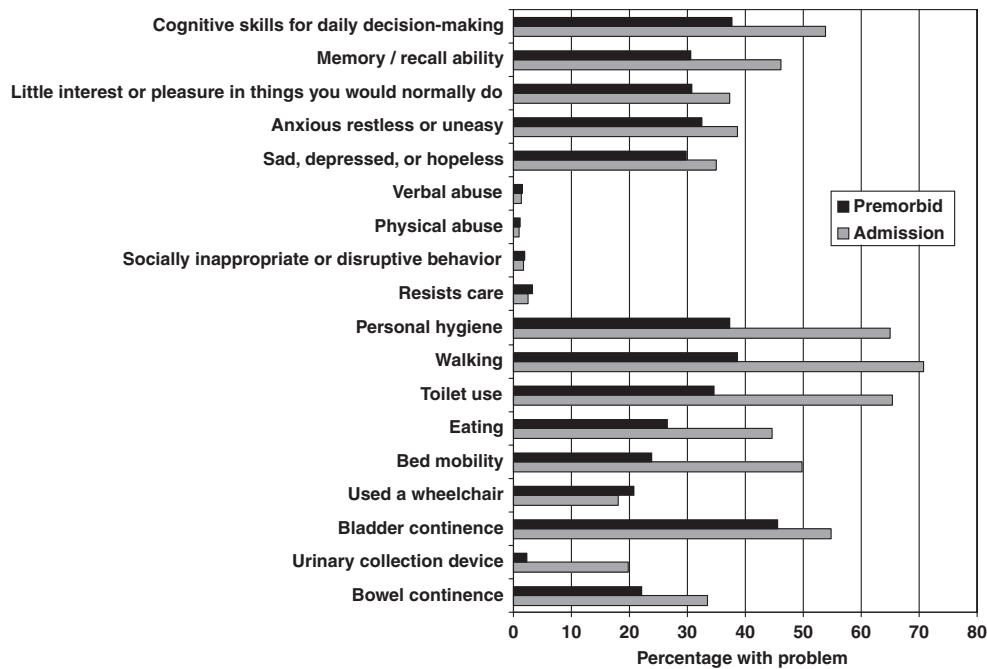
CI = confidence interval.

Figure 2. Proportions of patients with abnormal findings in the premorbid and admission observation periods.

lower in acute general medical patients than the other two groups in both the premorbid and admission observation periods. The change in frequency of deficits between the two observation periods was most marked in the orthopedic group.

Interrater Reliability

In total, 51 clinical items were evaluated. This comprised 18 items recorded at both premorbid and admission observation periods and an additional 12 premorbid only and 21 admission only items, producing a total of 69 individual data items for analysis. Of these 51 items, eight have dichotomous responses (usually yes/no), and 43 have ordinal responses. For the former, kappa values are reported for the dichotomous items and weighted kappa values for the ordinal items. Overall, interrater reliability was substantial ($\kappa > 0.6$) among the clinical items (Table 1). The average kappa was 0.61 for premorbid items and 0.66 for admission items. Of the 69 items tested, less than moderate agreement ($\kappa < 0.4$) was recorded for six (9%), moderate agreement ($\kappa = 0.41$–0.6) for 14 (20%), substantial agreement ($\kappa = 0.61$–0.8) for 40 (58%), and almost perfect agreement ($\kappa > 0.8$) for nine (13%).

Items of particular interest were those in which considerable assessor judgment is involved. These items relate to cognitive performance, mood, ADL and IADL performance, and pain. There was substantial interrater agreement for almost all items in these domains. Items that were recorded in both premorbid and admission periods performed equally well, with few exceptions (generally for low-incidence items).

DISCUSSION

The majority of items included in the instrument exhibited high frequencies across all sites and clinical settings, justifying their inclusion.

The prevalence of abnormalities in clinical items was high in the premorbid period, and the extent of change between the premorbid and admission periods was also high, indicating the importance of collecting premorbid information. Premorbid deficits are associated with a range of adverse outcomes in the acute care setting, underpinning the need for a robust method to record them.[11–14] High levels of deficits in the areas of cognitive function, performance of ADLs, mobility, and continence in both the premorbid and admission observation periods indicate that older hospital inpatients frequently have classic geriatric syndromes and reinforces the need to have adequate systems in place to identify, record, and respond appropriately.

In only a few items was the prevalence low. Most notable was the low prevalence of behavioral problems. This was in spite of a high prevalence of other observations suggesting the presence of delirium. During the period of observation, the majority of cases of delirium appeared to be of the hypoactive type.[15]

Assessors used in this study were necessarily inexperienced in the use of the instrument. Subsequent experience with large numbers of cases suggests that performance improves as assessors gain more experience with an instrument. The method of data recruitment, which involves multiple sources of information, might be expected to be associated with low levels of agreement. This same method of data collection is applied in other interRAI instruments, but usually this is with more-stable populations in home or long-term institutional care. Performance of the interRAI AC approaches performance of similar items in these other settings.

Particularly reassuring was the satisfactory performance of two groups of items. The first included the items relating to the premorbid period. These often involve an interview of family members—sometimes by telephone—and it might be anticipated that it would be associated with high levels of variation. Second, it might also be anticipated that items

requiring good observational skills on the part of assessors and their clinical colleagues would be associated with high error rates. These include the observation of cognitive function (including delirium symptoms), ADL function, and continence. Published studies indicate that the reporting of these items in medical records is poor.[16] In both cases, substantial or near-perfect interrater reliability was demonstrated.

The interRAI AC provides crucial information that might create the opportunity to adjust episode-based case-mix systems such as Diagnosis Related Groups to take into account functional and psychosocial parameters.

In some cases, collection of data is time consuming, but once it is collected and assembled in electronic form, there are many opportunities to improve the efficiency of data management and the quality of care. Studies are required (and planned) to examine the "return on investment" related to the use of this instrument. Comprehensive geriatric assessment is most effective when targeted at patients that meet particular criteria.[17] It is likely that this instrument will offer most utility when applied in designated geriatric units and in geriatric consultation. In other settings, it is possible that a screening strategy will be required to identify the most suitable patients.

CONCLUSION

After this study, the instrument development panel modified or deleted items that performed at less than the substantial level. It is anticipated that performance of the modified items will have improved in the final version, with the overall interrater reliability greater than in the trial version.

The final version of the instrument data set was released in 2006, but instrument development continues, including studies to validate the integrated diagnostic and risk screening tools against either clinical (criterion standard) assessments or "best of breed" existing screening tools and studies to validate the clinical assessment protocols against measured outcomes. Shortened versions designed to screen for patients requiring full assessment, including application in the emergency department, are currently being developed. Finally, algorithms to support case-mix classification and quality indicators, akin to those developed by interRAI for community and long-term care, are planned.

ACKNOWLEDGMENTS

REFERENCES

1. Philp I, Newton P, McKee KJ et al. Geriatric assessment in primary care: Formulating best practice. Br J Community Nurs 2001;6:290–295.
2. Reynolds T, Thornicroft G, Abas M et al. Camberwell Assessment of Need for the Elderly (CANE). Development, validity and reliability. Br J Psychiatry 2000, 444–452.
3. INTERRAI. Available at www.interRAI.org Accessed May 1, 2007.
4. Carpenter GI, Teare GF, Steel K et al. A new assessment for elders admitted to acute care: Reliability of the MDS-AC. Aging (Milano) 2001;13:316–330.
5. Morris JN, Fries BE, Morris SA. Scaling adls within the MDS. J Gerontol A Biol Sci Med Sci 1999;54A:M546–M553.
6. Morris JN, Fries BE, Mehr DR et al. Mds cognitive performance scale. J Gerontol 1994;49:M174–M182.
7. Fries BE, Simon SE, Morris JN et al. Pain in U.S. nursing homes: Validating a pain scale for the minimum data set. Gerontologist 2001;41:173–179.
8. Hirdes JP, Fries BE, Morris JN et al. Integrated health information systems based on the RAI/MDS series of instruments. Healthc Manage Forum 1999;12:30–40.
9. Fleiss J. Statistical Methods for Rates and Proportions, 2nd Ed. New York: John Wiley & Son, 1981.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
11. Alarcon T, Barcena A, Gonzalez-Montalvo JI et al. Factors predictive of outcome on admission to an acute geriatric ward. Age Ageing 1999;28:429–432.
12. Beech R, Withey C, Morris R. Understanding variations in lengths of stay between hospitals for fractured neck of femur patients and the potential consequences of reduced stay targets. J Public Health Med 1995;17:77–84.
13. Campion EW, Jette AM, Cleary PD et al. Hip fracture: A prospective study of hospital course, complications, and costs. J Gen Intern Med 1987;2:78–82.
14. Incalzi RA, Capparella O, Gemma A et al. The interaction between age and comorbidity contributes to predicting the mortality of geriatric patients in the acute-care hospital. J Intern Med 1997;24:291–298.
15. Inouye SK, Foreman MD, Mion LC et al. Nurses' recognition of delirium and its symptoms: Comparison of nurse and researcher ratings. Arch Intern Med 2001;161:2467–2473.
16. Jonsson PV, Finne-Soveri H, Jensdottir AB et al. Co-morbidity and functional limitation in older patients underreported in medical records in Nordic Acute Care Hospitals when compared with the MDS-AC instrument. Age Ageing 2006;35:434–438.
17. Winograd CH. Targeting strategies: An overview of criteria and outcomes. J Am Geriatr Soc 1991;39:25s–35s.