THE STANDARD ASSESSMENT OF GLOBAL ACTIVITIES IN THE ELDERLY (SAGE) SCALE: VALIDATION PROCESS OF A NEW TOOL FOR THE ASSESSMENT OF DISABILITY IN OLDER ADULTS.

A DESIGN THESIS
By

IRENE MARZONA, PharmD

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements for the Degree
Master of Science
McMaster University

MASTER OF SCIENCE (2011) in Health Research Methodology, Clinical Epidemiology Stream

McMaster University, Hamilton Ontario

TITLE: The Standard Assessment of Global activities in the Elderly (SAGE) scale: validation process of a new tool for the assessment of disability in older adults.

AUTHOR: Irene Marzona

SUPERVISOR: Dr. Koon Teo

NUMBER OF PAGES: VIII, 111

Acknowledgements

I would like to thank Dr Salim Yusuf and Dr Teo Koon for giving me the great opportunity and privilege to learn from their invaluable experience in Clinical Research and be part of the Health Research Methodology Program as a student.

My fellowship at the Population Health Research Institute wouldn't have been so rewarding without: Dr Martin O'Donnell and Jackie Bosch.

Thanks to Dr Mary Law, Dr Lehana Thabane and Dr George Heckman for the suggestions in this Thesis process.

To all the friends who welcomed me when I first came to Canada. John and Donna Bortolussi, Solange and Doug Trudel, Ingrid and Jay, Fulvia Baldassarre, all the members of the Famee Furlane Club of Hamilton plus all the colleagues of the PHRI (Jenn S, Jenn C, Amy, Edith, Tracy, Filomena, Linda, Donna, Stephany): THANK YOU !!! I've never felt alone.

This accomplishment, this step forward in building my knowledge is dedicated to whom I love most: *Mamma e Nonna*

Abstract

Background The possibility of ageing independently during the past 2 decades assumed a meaning which comprises different aspects. It has been recognized, in studies involving older adults but also by important Health Organizations (such as WHO), that disability could originate from different causes: physical limitations, external causes (such as personal assistance or building barriers), individual causes (lifestyle, behavior, positive attitude) and societal factors. Measuring the level of disability in a comprehensive way could help predict the amount of help and the best resources needed for older adults to cope with disability and remain independent as much as possible. The SAGE scale has been developed to be a complete and easy to use tool to measure independence in older adults.

Objectives The aim of this thesis is to describe the methodology and the design of a study and to assess the validity and reliability of the SAGE scale.

Design SAGE validation will be measured in a cross sectional study, involving 240 older adults conveniently sampled from 3 different facilities in the Hamilton area. Community dwelling older adults, patients recovering from a stroke and subjects living in a nursing home, will be assessed at one point in time with the new tool (the SAGE scale) together with four widely used scales to assess cognitive abilities (Montreal Cognitive Assessment), functional abilities (The Franchay Activity Index), the modified-Rankin scale and the CSHA Clinical

Frailty Scale. This thesis will describe the processes through which assess the content, construct and criterion validity. The hierarchical sequence of items will also be investigated as well as specificity and sensitivity of the new tool.

Conclusion The development of the SAGE was motivated by the recent need for an instrument able to capture all the activities that are important for the elderly to be able to age with dignity and independence. The results of this study, if positive, will be useful for further investigation of the SAGE, as a screening tool to recognize and detect early loss of independence in this group of individuals.

Table of Contents .....Page

Chapter 4:The Validation process: an overview

Chapter 5: Design of the SAGE validation study

Chapter 6: Assessing SAGE validity

Chapter 7: Conclusions and Future Directions

Chapter 1 Introduction

Introduction

In the past decades, the importance of detecting the degree of disability in the elderly has been recognized. With the expansion of the life span, the early detection of functional and cognitive decline can help the assessment of the degree of disability and therefore the level of independence an individual can achieve. In this context, a tool able to detect the loss of early independence can therefore help in properly predict the resources that will be necessary for the individual to function and continue to live in the community. Such an instrument should be able to capture the activities and the abilities that are most important to an individual to live and participate in the society independently. The aim of this thesis is to describe the development of a new scale for the comprehensive measurement of functional and cognitive abilities in the elderly and to examine how to properly assess the validity of this scale.

1.1 Functional status

Functional ability is the capacity to carry out daily activities at a personal and at a community level[1]. Functional independence has been recognized to be an important indicator of a cognitively healthy status in assessing a subject's well being. Verbrugge described a model in which he indicated clearly how functional limitations such as not being able to ambulate, reach and climb stairs, could lead to more important disabilities, for example taking personal

care, managing the household, having a job or hobbies. Disability is defined as the inability to perform activities of daily life[1]. As the human life span increases and the proportion of older adults grows, it is very important to assess the level of functional ability to predict the grade of disability in the short and long term. As an example, in a longitudinal study performed on more than one thousand community living people (aged 72 years or more), participants were assigned at entry into 3 categories: independent without difficulty, independent with difficulty and dependent in performing basic activity of daily living (BADL) such as bathing, toileting, walking and dressing. Most participants in this study reported only one or two difficulties, usually with bathing and dressing. Over a 3 year period participants who had difficulty with BADL were more likely to develop a dependence with BADL (RR 1.7 [95% CI 1.3-2.2]), and patients in the dependent group had a significantly (p=0.017) higher probability of entering a “skilled” nursing facility (probability= 0.35), compared to patients independent with some difficulty (probability= 0.22) and to patients independent without difficulty (probability= 0.10)[2]. These findings suggest that difficulty and dependence are complementary information on the functional ability in a patient, and the loss in function leads to dependence and admission to nursing care facilities. The early assessment and evaluation of patients with disability could therefore help predict patients’ needs for help in performing basic activities or the need for recovery in nursing care facilities.

Moreover it's disability and not age that lead to the loss of independence in subjects living in the community.

1.2 Participation

Being able to be involved in the community and societal activities is perceived to be as important as being able to self care by older adults. The World Health Organization in the 1980's recognized that the ability to participate in societal activities is one of the direct consequences of health issues[3]. During the ageing process the inability to perform activities at a societal level is the first manifestation of important loss of independence[4]. In a cross sectional study comparing patients who were not disabled and aged between 65 and 85, participants 85 years old scored significantly lower in the personal care, communication, housing and mobility items and in the interpersonal relationship, community life and leisure item of the social domain in the LIFE-H scale short version 3.1[5]. Participation restriction was found to be associated with health impairments such as dizziness and weakness of limbs, with health conditions (heart, chest problems), anxiety, depression and cognitive impairment in a logistic regression age-gender adjusted model on more than 7000 participants, aged 50 years or older, of the North Staffordshire Osteoarthritis project (NorStOP). Marital status[5,6], unemployment and poor perceived adequacy of income were also associated with participation restriction [6]. Although participation restriction depends on physical and social individual characteristics, environment and perceived barriers were also

reported to be related to a person's ability to participate in community activities[4]. It is therefore important also to evaluate the level of involvement of an older individual in the community and in the society to evaluate his overall level of independence.

1.3 Cognitive decline

Age related cognitive decline is an important health issue with an increasing prevalence as the population ages. Cognitive decline leads to significant functional loss and is a major component of age related deterioration[7]. Dementia, a major contributor to incapacitation and institutionalization rises rapidly doubling every 5 years after age 65 years[8]. Cognitive decline and dementia are important health issues which also consume enormous amounts of financial resources. These diseases, have been shown to be associated with a number of cardiovascular diseases. Hypertension is shown to be correlated with increased risk for cognitive vascular impairment due to overt stroke or small vessel disease in the brain which in most cases constitutes a silent form of cerebral ischemic disease[9,12]. For example, in the Sys-Eur trial (n= 2418), patients aged 60 years and over with isolated hypertension were randomized to a treatment with nitrendipine (calcium channel blocker) or placebo. During 5 years of follow up, patients were evaluated at each follow up visit with the Mini Mental State Examination (MMSE). The MMSE is a widely used test to assess cognitive function over a broad spectrum of 30 items in which by convention demented patients score 23 points or less. After

a median follow up of 2 years in the study, a total of 32 cases of dementia were recorded, 23 of which were due to Alzheimer's disease. There were 21 cases in the placebo group and 11 in the active treatment group showing a reduction in the cases of dementia of 50% (95% CI 0-76; p= 0.05) on intention to treat analysis[10]. Although high blood pressure is the major cause of large vessels stroke, low blood pressure through a hypoperfusion mechanism can also be responsible for small vessels infarct mainly in the deep white matter[9,13]. Incident lacunar infarcts were also found to be significantly correlated with atherosclerotic phenomena such as carotid intima thickness with OR 1.27 (95% CI 1.01; 1.61) and presence of carotid plaques with OR 1.17 (95% 1.00; 1.37)[11]. Cognitive decline and dementia were also found to be related to unrecognized myocardial infarction (MI) in men in the Rotterdam study involving 6347 non demented participants at baseline who underwent electrocardiography between 1990 and 1993. In the Rotterdam Scan Substudy 436 non demented participants of the first study underwent MRI scan until 2005, cognitive testing was also performed in 3 follow up visits. In men, unrecognized MI were found to be associated with an increased risk of dementia compared to men without MI, HR 2.23 (95%CI 1.24; 4.01) with the analytic model fully adjusted for all important covariates. In the same prospective study, both recognized and unrecognized MI were associated with a higher risk of having white matter lesions, OR 3.50 (95% 1.54; 7.96) and OR 7.19 (95% CI 1.17-44.07) [14]. It has also been reported that diabetes mellitus is

also associated with an increased risk of dementia (RR 1.9, 95%CI 1.3; 2.8)[15]. Increased risk of dementia[15] or mild cognitive[16] decline was found to be proportional to the duration and the severity of the disease. Atrial Fibrillation (AF) is also a consistent risk factor for stroke and therefore for cognitive decline.

What is clear from this brief review of the available literature is that there are modifiable risk factors which increase, unless uncontrolled, the risk of dementia and cognitive impairment. Therefore it is desirable that more strict control and the early assessment of those risk factors be available to assess the progression of the cognitive status.

1.4 Frailty

Frailty has been defined as a complex status of vulnerability due to impairments in multiple systems separated from the normal ageing process that result in multiple adverse outcomes. Frailty is related to chronic diseases, but also from cognitive function and mood[17]. Recently it has been recognized that also poor social condition and support are part of the frailty syndrome[18]. A tool able to screen subjects for early frailty has not yet been created, nowadays there are some rule-based definitions to help clinicians assessing frail older adults**.**

1.5 Successful ageing concept

The first investigators to introduce the concept of successful ageing were Rowe and Kahn, who separated the effects of the diseases from the process of the normal ageing[19]. According to their concept, subjects who age successfully would maintain high psychological, cognitive functioning while avoiding clear disease and disability. In their work they also urged further research to understand the pathway of age- related transition in loss of function between successful to usual, usual to diseased, diseased to impaired function. Rowe and Kahn also pointed for the first time that psycho-social support and more importantly, autonomy were important determinants in reversing the process of loss in function[19].

According to a 2006 exhaustive review of observational prospective studies to identify predictors of successful ageing, the average number of component predictors used to define "successful ageing" was 2.6 (SD 0.4). Physical functioning/disability were the most frequently cited as components of successful/unsuccessful agers, mostly measuring the ability to carry on activities of daily living (ADL) and less frequently instrumental activity of daily living (IADL). The next frequent component was found to be cognitive function, either measured through cognitive testing administration (i.e. Mini Mental State Administration) or through self reported memory function. Despite the high variability in the choice of the items used to individualize the successful agers, some trends in the predictors among the studies considered in the review were evident. Younger age, absence of hearing problems or

absence of arthritis, were consistently related with successful ageing across more than 75% of the studies included in the systematic review. Higher exercise or physical activity levels, better self related health, lower blood pressure levels, global cognitive function and absence of depression were found to be related to successful ageing in more than 50% of the studies considered[20].

When interviewed on the meaning of "successful ageing" and the necessary components of successful ageing, based on in their personal opinion, older adults gave a different prospective. In a study involving 12 focus groups, each formed by 6 subjects all community dwelling and independently living aged between 60 to 99 years, 4 major themes associated with successful ageing were identified: attitude and adaptation, security/stability, health/wellness and engagement in stimulating and meaningful activities. Throughout the focus groups, the ability to accept change in relation to the normal ageing process in a positive, optimistic way was considered fundamental to successful ageing: some participants didn't distinguish between successful ageing and successful living. Living in a secure environment and financial stability were mentioned as the second important characteristics of successful ageing. Social support by a spouse or companion, family and friends was also associated with security and stability. Regardless of the form of such relationships, their presence was perceived to contribute to the participant's successful ageing by providing emotional and instrumental support. Being physically healthy was also

reported to be a contributor to successful ageing, although in some cases patients stated also that in case of illness, optimistic attitude could lead to a successful and positive ageing process. Participation in social activity was the last theme identified by participants, which was associated with intellectually challenging activities, enjoyment, have a purpose and remaining interested in activities that were important at a younger age[21]. In two large prospective cohort studies this trend was confirmed: 970 patients recruited from the Kame study and 1173 patients recruited from the Adults Changing in Thoughts study of participants 65 years or older, revealed the multidimensionality of the "successful ageing" concept. These cohorts of patients were interviewed on their personal interpretation of the "successful ageing" construct, by choosing between lists of attributes retrieved from the literature on the topic. Among the 20 attributes, both group of patients focused their attention on 13 in particular. As noted in the previous study, the respondents' physical health, independent functioning and active engagement with social activity were the attributes that best explained the concept of successful ageing[22].

1.6 Ageing in Place

Older adults should be able to age successfully in their own environment, choosing to live in their community as long as possible and having access to services that will support this choice[23]. This ability is called "ageing in place" and in the last twenty years different approaches have been suggested to help and support older adults live in their own environment by creating

communities and practices “elderly friendly”[24,25]. This approach has already been taken in consideration in Government policies of different Countries such as United States, Canada, Australia and in Europe France and England. Ageing in place is believed to be important in increasing the quality of life and dignity of older adults and represents also an opportunity to decrease the costs due to early institutionalization of this group of citizens.

Conclusion

The possibility of ageing independently (or ageing successfully) during the past years assumed a meaning which comprises different aspects. It has been recognized not only in studies involving older adults but also from important Health Organizations that disability could originate from different causes. As Verbrugge outlined in his model[1], not only physical limitations but also external causes (such as personal assistance or building barriers), individual (Lifestyle, behavior, positive attitude) and societal factors play a role in the process that leads to the inability of being independent in the community. Measuring the level of disability in a comprehensive way could help predict the amount of help and the best resources needed. Secondly, it is very important recognizing that dignity related to independence plays a major role in the perception of the disability level which is often based on older adults’ self reported information.

Different scales and indexes were developed to capture the self reported disability level, usually these scales were conceptualized on a population characterized by a particular disease. Some of these constructs were also validated in older adults, in the next section I will review all the scales validated in the elderly, dividing them accordingly to the type of activities measured.

Chapter 2 Background on available scales

Introduction

Disability in the elderly is strongly associated with chronic clinical conditions such as foot problems, cognitive impairment, arthritis, vision and coronary heart disease. In the Canadian Study of Health and Ageing (CHSA), these factors were found to be strongly associated with loss of independence in performing self care and instrumental activities[26]. In low and middle income countries the major contribution is given by dementia, but also chronic conditions such as the above mentioned, were found to be strongly associated with disability[27]. Disability in the elderly will be an increasing burden as the human life span increases, especially in women and in developing countries[28]. Early diagnosis of disability in the elderly, is important to help older adults adopting adaptive behaviors; the aim is to cope with disability and remain independent as longer as possible.

Measuring health status and disability was, over the last few decades, related to the level of illness and based on the use of diagnostic tests. Measurements of fitness for work or recovery after an injury were used to assess and standardize ratings on patient's abilities. Global impairment and disability is not only related

to the physical capacity of a patient but is also associated with different component of a subject's level of independence: environment, social support and patient's subjective attitude. Moreover the purpose of rehabilitation is to reintegrate patients into their original environment as soon as possible. Physical impairment alone cannot be sufficient to predict a successful return to active living.

Scales have been conceptualized to cover also activities that patients could or did perform at their level of physical capacity. This kind of measurement was defined as "functional disability", characteristic were scales measuring the level of difficulty an individual can have in performing a list of activities called: Activities of Daily Living.

2.1. Activities of daily living (ADL)

The term Activities of Daily Living (ADL) refers to the basic tasks of everyday life such as dressing, bathing, eating and transferring independently. People unable to carry out those activities require help from a caregiver or by using a mechanical tool. Information on the ability to perform ADL is related to the level of self care and independent living of the patients. The first scale on assessment of ADL was created by Katz in 1957 and included 6 activities: bathing, dressing, getting from and to the toilet, transfer from bed to chair, feeding and continence[29]. Katz and colleagues observed that in the recovering patients the restoration of independence went through three stages: first was

the independence in feeding and continence, second was transferring from bed to chair and going to the toilet, last and with the complete recovery were bathing and dressing. The regaining of functional ability in disabled patients, according to Katz et al, reflected the pattern of child growth and development, hypothesizing that as there is an orderly pattern of development of function in childhood there is also an ordered regression in the ageing process.

This hypothesis was confirmed in national surveys conducted in the United States around 1980 in studies with a large variability in the sizes of the elderly population with ADL disabilities. In non institutionalized persons aged 65 years or older, higher proportions of subjects needing help for bathing were found across all the surveys considered, followed by dressing, transferring, and toileting while eating was the least difficult activity to perform. The same trend of difficulty (with higher proportion of patients needing help) was observed also among institutionalized patients[30]. Most recently longitudinal studies confirmed that in both man and woman with median age 78 years, age at onset of disability was earlier for bathing, followed by mobility, toileting, dressing, transfer from bed, transfer from chair and feeding. When adjusted for age, comparison between man and women showed a higher risk in disability for bathing (RR 1.6; 95% CI 1.3-1.9) and getting from and to the toilet (RR 1.7; 95%CI 1.2-2.5) for women compared to men[31].

ADL SCALES AVAILABLE

The literature reports an infinite number of scales and indexes to measure abilities in self care tasks, for patients with different chronic illnesses. Therefore a comprehensive review is not considered suitable within this thesis. In table 1, a list of the most widely used ADL scales validated in older adults is summarized, with the type of validation methodology used and the results reported.

The Katz' index for ADL

This scale was developed from the observation of a large cohort of patients who suffered from the fracture of the hip and by rating the subjects on the basis of three different descriptions of the patient's ability in performing the following activities independently: bathing, dressing, getting from and to the toilet, transfer from bed to chair, feeding and continence[26]. The scale was validated in a wide range of patients and with different stages of disability. In the elderly, the scale demonstrated a good predictive value for patients rated as independent or dependent: 32 patients of the 44 rated as independent were living at home a year later while 54% of those rated as dependent died[32]. Up to 2007[33] the Katz scale was widely used in the evaluation of functional outcome in the elderly, although no proper validity and reliability evaluations could be retrieved from the literature. A recent study on Turkish, Moroccan and Dutch elderly people showed that the Katz index was strongly associated with the Organization for Economic Co-Operation and Development scale (OECD) a long term limitations in mobility indicator (Pearson's r= 0.64) and with 4 physical functioning items of

Short Form Health Survey (Pearson's r= -0.60). Moderate correlation were found with a measure of depressive symptoms (Centre for Epidemiological Studies- depression Scale) and with chronic conditions recorded at baseline. Although concurrent validity results of this study are positive, their generalizability to all ethnic groups have yet to be shown[34].

The Barthel Index

The Barthel Index (BI) was originally developed to assess the functionality and independence of individuals and to indicate the grade of nursing help needed in patients hospitalized for neuromuscular or musculo-skeletal diseases. It was also used for rehabilitation purposes: to predict the duration of disability and to estimate prognosis. The index was administered at the beginning of a rehabilitation treatment, during the rehabilitation process and at the time of maximum benefit.

The index was first published in 1965[35] and was composed of 10 different domains: feeding, moving from wheelchair and back, personal toileting, getting on and off the toilet, self bathing, walking on a level surface, ascending and descending stairs, dressing, controlling bowels, controlling bladder. The authors fully described levels of independence in performing the activities listed in the index by assigning scores ranging from 0 to a maximum of 15, a score of 0 was given in the case the patients couldn't meet the defined criterion for not needing some kind of help (points) or being independent (10 or 15 points). The overall

score could range from 0 to 100 and the scores reflect the amount of time and assistance the patient needs, while the maximum score defines completely independent subjects. The authors predicted that a score of 100 did not necessarily mean the patients could live alone but that they could take care of themselves.

Despite the population from which the Barthel Index originated, in 1967 Wylie assessed the validity of the index in 1025 patients admitted after a stroke episode. Thirty seven percent of patients with an initial score of 0 to 15 points in the Barthel Index, died over an average follow up period of 47 months and 77 % of patients who initially scored from 60 to 100 had improved when discharged suggesting the lower the initial Barthel Index score the less the probability of the individuals would be discharged as clinically improved[36,37].

A more accurate analysis on the property of the Barthel Index in 396 hospitalized older patients showed that despite the wide use of the instrument in the acute setting, the scale is not unidimensional nor has interval properties. This means that the distance between scoring supposed to give an indication of the disability/ability of the patients is not a constant interval and the dimension measured is not only the level of ability[38].

Lawton- Brody Physical Self Maintenance Scale

The ADL version of the Lawton- Brody (the Physical Self Maintenance Scale-PSMS) scale was published in 1969 and focused on observable behaviors of

the subjects. It comprises 6 items: toileting (which includes also information on continence), feeding, dressing, grooming, physical ambulation, bathing. Responses are on a five point scale ranging from total independence to total dependence. The characteristics of the subjects included in the validation process, suggested to the Authors to broaden the spectrum of activities that was possible not only for nursing home patients but also for community dwelling. The IADL (Instrumental activities of Daily Living) list included: ability to use the phone, shopping, food preparation, housekeeping, laundry, mode of transportation, responsible in taking own medications, ability to handle finances. The property of IADL scales will be discussed later, but in this context it is important to mention that in the validation process of the Lawton-Brody scale, a higher correlation (r= 0.61) was found between PSMS and IADL, which suggested a hierarchical association between the two batteries of activities. As expected the PSMS correlates less strongly with a measure of cognition (r= 0.38) or a measure of behavior adaptation (r= 0.31)[39]. No further investigation on validity of PSMS in an older adult population has been retrieved from the literature.

Functional Independence Measure (FIM)

The Functional Independence Measure was developed to measure disability in 18 basic items of daily living, and was divided into two subscales: motor and

cognitive. The Motor items included eating, grooming, bathing, dressing upper body, dressing lower body, toileting, bladder management, bowel management, bed- chair/ wheelchair transfers, toilet transfer, tub-shower transfers, walk/ wheelchair locomotion, climbing stairs. The cognitive domain consisted of 5 items: comprehension, expression, social interaction, problem solving and memory. Items were scored on a 7-point scale in an ordinal way from most dependent to independent. FIM was designated to be used in the older adults and measure the level of assistance required. Although the cognitive FIM subscale demonstrated a poor predictive value in patients with progressive multiple sclerosis, the psychometric properties of the motor scale were found to be similar to the Barthel Index in stroke and neurologic patients[40,41]. Pollak et al evaluated the validity and reliability of the FIM by administrating the tool to 49 subjects, 80 years and older, twice. The content validity (the extent to which a tool measures all the aspects of disability) of FIM was measured with Rasch analysis, for the motor subscale bladder and bowel incontinence and resulted in misfitting of the items due principally to operational definitions of these items in the FIM. However reliability was very good for both subscales (motor ICC=0.9 cognitive ICC=0.8)[42].

Conlcusion on ADL scales in the elderly

Measuring disability in the elderly has become important not only in predicting outcome from rehabilitation but also in determining the burden of the disability in terms of assistance needed and required. Therefore ADL scales are used

mostly in assisted environments or in acute care settings and provide information useful in planning follow up and the progression of the dependence status in the subjects.

Table 1.VALIDATED ADL SCALES in ELDERLY

Name, Year, Reference	Population	Items	Validation Method	Validaty-Reliability Results	Methodological consideration
Katz, 1963[29], **1984**[32] **2006**[34]	100 aged patients 993 subjects (55-74 yo): Dutch, Moroccan, Turkish	6 items 6 items	Guttman scale analysis, Predicitve validity Internal consistency across ethnic groups. Concurrent validity against SF-36 PF, OECD, CES-D	Good reliablility and scalability After 1 year: 75 % independent Subj. were living at home; 54% of dependent had died Internal Consistency (0.84-0,94), good correlation with SF-36 (-0.60) and OECD (0.64).	Proper validation in different ethinc groups.
Barthel, 1967 [35], **2008** [38]	396 Acute medical older patients	BI 10 items, 3 responses MBI 10 items 5 responses	Assessment with BI and modified BI at admission and discharge. Rasch analysis	The Rasch analysis showed that BI is not a unidimensional measure of ADL Continence items showed misfitting with the model	Unidimensionality not proven. Good discrimination power. Rescoring of the misfitting items was necessary, therefore not suitable for primary care and acute care use.
Lawton-Brody PSMS 1969 [39]	265 community dwelling or living in nursing home subjects aged 60 year or more	6 items, 5 point response scale	Construct Validity : Physical Classification (PC), Mental Status Questionnaire (MSQ), Behavior Adjustment	Reproducibility coefficient: 0.96 Pearson's correlation: PSMS-PC: 0.38 LB IADL: 0.61 BA: 0.38	Further validation studies not retrieved.

			(BA), Lawton-Brody IADL		
FIM 1996[42]	49 patients from Skilled Nursing Home, Sheltered Care or Independent Community aged 80 years and older	18 items divided in 2 domains: motor and cognitive	Content validity: Rash analysis Test retest reliability by subscale	Similar calibration noted between for upper body dressing, toilet transfers, and bed/chair/wheelch air transfers. Test-retest reliability for 45 repeated FIM assessments for the motor, (ICC = 0.9). cognitive (ICC =0 .8)	Enviromental factor and operational definitions influenced FIM ratings. Useful in the elderly population

2.2 Instrumental Activities of Daily Living (IADL)

In order to be able to broaden the evaluation in community dwelling older adults, a further step in the evolution of the activities of daily living scales had to be made. ADL scales underestimate the proportion of the elderly needing full assistance because they do not include activities which require adaptation to the environment. Therefore the Instrumental Activity of Daily Living (IADL) scales were designed. The IADL scales extend the ADL ones to include tasks which require a higher level of complexity. IADL scales are used in populations that are less severely impaired and therefore require less assistance in performing activities which are considered basic, but also those that are not considered fundamental but still important for the individuals to be able to remain independent. IADL are tasks that may confer the ability of the individuals

to age and adapt independently in the community. They include financial management, medication management, telephone use, shopping, preparing meals, managing a household and transportation.

It could be supposed that community dwelling older adults have less difficulty in performing IADL activities than nursing home patients. In epidemiological studies of community dwelling and assisted care patients, an extended version of the ADL scale by Katz consisting of instrumental activities of daily living items was validated. The functions analyzed were two IADL (shopping and transportation) and five ADL (bathing, dressing, transferring and feeding). In more than 3000 individuals dependence in IADL and ADL increased with age, and onset of dependence in IADL activities started at an earlier age. Overall, it was also observed that individuals who were dependent in IADL were more likely to become dependent also in ADL like activities. Risk of death and hospitalization within one year was directly associated with the grade of dependency in IADL activities[43].

In a 25,470 population sample obtained from the Phase II of the National Health Interview Survey on Disability, information on ADL and IADL activities were obtained by administering to adults 18 years or older an extended scale including 5 ADL and 9 IADL (grocery shopping, getting to places, doing light housework, preparing meals, getting outside, walking, managing money, managing medications, using the phone). Guttman's method was used to evaluate the scalability, which measure how reliably persons and items can be

hierarchically ranked ordered from low to high levels of the scale. Item Response Theory (IRT) tests were used to assess the fit of the item in a probabilistic model depending on the level of trait and the difficulty of the items. ADL activities resulted to be biased by age (Coefficient of Scalability= 0.77), reaching a high level (CS= 0.93) at 85 years or older, while IADL scalability scores increased with age as well but remained lower than ADL (CS= 0.58), rising to 0.74 at 85 years or older. Therefore including an extended scale which comprises IADL items tested in a large population with various grade of disability seems to be more useful than the ADL scale alone. The Rash analysis of the fitting model also demonstrated unidimensionality of the extended version of the IADL/ADL which means it could be useful to detect various levels of disability in a broad range of ages[44].

In table 2 are summarized the IADL scale validated in older adults population available in the literature.

IADL SCALES VALIDATED IN THE ELDERLY

THE LAWTON – BRODY IADL SCALE

To measure less severe grades of disability, Lawton and Brody introduced a new index based on individuals needing help over eight instrumental activities of daily living[39]. Recognizing that independence should be measured by taking into account different tasks, with different sex-related and-age related competences, the authors produced a list of IADL which included the ability to

use the phone, shopping, food preparation, housekeeping, laundry, using transportation, managing own medication and ability to handle finances. Responses ranged on 3 to 5 point scale indicating how independent were the subject in carrying out the activities.

The Lawton and Brody scale for IADL showed a slightly different scaling between man and women, due to the fact that some items such as doing laundry, preparing meals and housekeeping were not scored in man. Pearson's correlation showed a significant fair correlation with the measure of cognition (r= 0.48) and with the Physical self- maintainance Scale (r= 0.61). The Lawton Brody scale is widely used in the assessment of disability in the elderly because it is easy to administer in 10-15 minutes. Although it is considered the best tool for the assessment of dependence in everyday life tasks, the self reported or proxy reported responses could lead to an under estimation or over estimation of the real status of the patient.

THE FRENCHAY ACTIVITY INDEX

The Frenchay Activity Index (FAI) was developed in 1983 primarily with the goal of assessing and managing rehabilitation in post stroke patients[45]. For the first time assessment of independence included not only tasks that could be carried on at home, but also at a "societal level". Therefore with the FAI a further step is taken in recognizing what is important to function independently in the rehabilitating patients: being able to have a social role is equally fundamental

than being able to self care. The FAI includes 15 items and it requires the self reported frequency with which the patient carries on the activities listed within the previous 3 months (preparing main meals, washing up, washing clothes, light housework, heavy housework, local shopping, social occasion, walking outside for more than 15 minutes, actively pursuing hobby, driving car or taking a bus) and within the last 6 months (travel or taking car rides, household or car maintenance, reading books, gainful work).

The FAI has been validated in different cohorts of stroke patients[46] revealing a high correlation with ADL measure, and measures of depression and IQ. Other validation studies recently evaluated the assessment of South Korean community-dwelling subjects aged 60 years or older. The reliability was found to be acceptable (Chronbach's $\alpha = 0.796$), the discriminative validity was assessed comparing the FAI scores in subjects using assistive device versus non users and there was a significant mean difference in the average score between users and non users of -4.67 ($p < 0,001$). Factor analysis used to determine content validity resulted in the individualization of 4 main factors: domestic tasks, outdoor work, leisure, hobbies[47]. The present study provides further information on the reliability of the FAI in a cohort of older adults with no clinically important characteristics, although a proper construct and discriminative validation in a generalizable ethnic group is still required to be able to conclude that the FAI is a valid instrument in assessing disabilities in community dwelling subjects.

THE ASSESSMENT OF LIVING SKILLS AND RESOURCES (ALSAR)

The Assessment of living skills and resources scale is composed of 11 items and resources such as: telephoning, reading, leisure, medication management, money management, transportation, shopping, meal preparation, laundering, housekeeping and house maintenance. Responses are ordered in five description of accomplishment of the task, and relative information on the availability of the external resource the patients need to complete the tasks. Both skills and resources are scored, giving a maximum of 4 point for each item. The ALSAR scale was administered to elderly veterans in a home care program with the aim to predict change towards living skills, hospitalization and institutionalization in a 6 months period.

In a recent comprehensive analysis the ALSAR scale was administered at 160 older adults[48], it's not mentioned whether community dwelling or chronically ill. Factor analysis was used to individuate the construct underlying the scale and Rasch analysis to evaluate the quality and the fitting of the items. Factor analysis demonstrated the unidimensionality of the scale, while Rasch analysis showed that home maintenance, housekeeping, shopping, using transportation, doing laundry and preparing meals were the most difficult items endorsed. Using telephone and reading turned out to be misfitting with respect to the model, probably due to vision impairment or difficulty in reaching the phone for some subjects. The ALSAR also demonstrated no differential items functioning (DIF) between the genders, showing that the scale is generalizable across

sexes. Overall the ALSAR scale resulted to be a valid instrument to measure the level of disability in the elderly although the scoring system in the study cited previously has changed in favor of one more clear and incremental with different level of skills and resources.

THE KOHLMAN EVALUATION OF LIVING SKILLS (KELS)

The Kohlman Evaluation of living skills (KELS) was developed for psychiatric outpatients. The KELS consists of 17 daily living skills grouped into 5 categories: self-care, safety and health, money management, transportation and telephone use, work and leisure. Scores are divided into two categories: independent and needs help. There are specific criteria defining the two categories and each criteria is scored differently and then computed. People achieving scores between 0-5.5 are considered able to live independently in the community and those with scores between 6 and 17 are considered to be in need of assistance in order to live independently in the community.

The KELS has already been validated in older adults in acute care hospital settings, showing a good inter-rater reliability agreement ranging from 0.74 and 0.94. Particularly interesting was a study on 92 older Israeli adults: 34 were living in the community, 44 were living in sheltered housing facilities and 14 were individuals living in the community but attending daycare services. The aim of the study was to determine construct and criterion validity of the KELS in older adults with different levels of dependence. KELS was proven to be a valid

tool in discriminating between subjects living in the community in sheltered housing or using day care facilities: 91% of the subjects living in the community turned out to be independent and 9% were at borderline. The proportions relative to subjects living in the sheltered houses was more variable with 64% found to be independent, 9 % were at borderline dependence and 27% needed assistance. Among the group of people attending daycare, 74% needed assistance and only 26% were considered independent.

KELS score was also correlated to the scores obtained with the Mini Mental State Examination (MMSE) which is a widely used test for cognitive status, with the FIM and with the Routine Task Inventory (RTI) which assessed the cognitive functional level in daily tasks. Spearman's Correlation were significantly high between KELS and MMSE (r=-0.757), between KELS and FIM (r= -0.707) and between KELS and RTI (r= -0.895)[49].

In a recent study criterion validity of KELS was also assessed on 200 community dwelling subjects aged 65 years or older. To evaluate cognitive, affective, executive and functional level validated and widely used scales were used to determine whether KELS could be a valid tool to screen patients able to live independently and safely in the community. The indexes used were the MMSE for cognition, the Geriatric Depression Scale (GDS), the modified Physical Performance Test (mPPT), the Knee Extensor Break Test, Executive Cognitive Test (EXIT25), executive clock drawing test (CLOX) and 8-foot walking test. KELS highly correlated with measures of executive function

(EXIT25 r= 0.705, CLOX r= -0.629) mildly correlated with measures of cognitive function (MMSE r= 0.508) affect (GDS r= 0.318) and physical function (mPPT r= 0.472), but it did not correlated with the Knee Extensor Break test[50].

THE LATE LIFE FUNCTION AND DISABILITY INSTRUMENT (LL-FDI)

The Late Life FDI is an instrument developed at the Boston University, specifically to measure function and disability in community-dwelling older adults. The LLFDI contains items that are representation of functional limitation (inability to perform physical tasks, subdivided in three dimensions such as upper extremities, basic lower extremities and advanced lower extremities) and disability (inability to be part of life tasks and have a social role). The instrument measures the self reported difficulty a patient encounters in performing 32 enlisted functional activities by choosing between none, a little, some, quite a lot and cannot do. The disability component is measured through the self reported limitation of taking part in 16 major tasks by choosing between the following answers: not at all, a little, somewhat, a lot or completely. Some questions are also phrased to be answered in terms of frequency and in this case the possible answers are: very often, often, once in a while, almost never and never. Both components in different analysis demonstrated good test retest reliability: for the 32 items functional dimension ranging between 0.91 and 0.98[51] and for the 16 item disability component 0.68-0.82[52].

The instrument has been validated against two test of functional ability: The 400 meter Self paced walk (400-m W) and the short physical performance battery (SPPB). The first one consists of asking patients to walk for 400 meters at their own pace and measuring whether the 400 meters were completed or not, time to completion and heart rate. SPPB is a performance test assessing lower extremities function using measures of gait speed, standing balance and lower extremities strength. As expected a good correlation was found between the functional component of LL-FDI and the 400–m W ($r= 0.69$ $p< 0.001$), while no predictive correlation was found on limitation or frequency of disability component items ($r= 0.44$ $p<0.001$ and $r=0.2$ $p> 0.004$ respectively). Similarly a good correlation was found between LL-FDI and SPPB regarding the function component ($r= 0.65$ $p<0.001$), but no correlation was evident with limitation indicators ($r= 0.37$ $p< 0.001$) or frequency in endorsing social and life activities ($r=0.16$ $p> 0.004$)[53].

CONCLUSION ON IADL SCALES

Scales and indexes which measure function at the instrumental level usually include activities that can be carried in a home environment. It is not always required to indicate if the help of a person or a tool is needed to be able to complete the activity. IADL scales care intended to be used in a broader population which still lives at home and are able to be independent in their usual environment.

Table 2. IADL VALIDATED SCALES IN THE ELDERLY

Name, Year, Reference	Population	Items	Validation Method	Validaty-Reliability Results	Methodological consideration
Lawton-Brody IADL, 1969 [39]	265 community dwelling or living in nursing home subjects aged 60 year or more	8 items, 5 point response scale	Construct Validity : Physical Classification (PC), Mental Status Questionnaire (MSQ), Behavior Adjustment (BA), Lawton-Brody IADL	Reproducibility coefficient: 0.96 Pearson's correlation: IADL-PC: 040. IADL-MSQ: 0.48 IADL-PSMS: 0.61	Further validation studies not retrieved.
Franchay Activity Index, 1983[45], **2009**[47]	770 community dwelling subject, average age 62 years	15 items	Internal consistency Content validity Construct validity Factor analysis	Chronbach's α= 0.796 Assistive device users vs non users - 4.67 FAI score (< 0.001) Factor analysis: 4 factors	Further validation in community dwelling and nursing home patient generializable cross culturally.

ALSAR (1991) 2009[48]	160 older adult (total of 290 ratings)	11 items	Rasch Analysis Factor Analysis Reliability	Factor analysis: unidomensionality Rasch analysis: telephone use and reading are misfitting items (due to inability of some subjects to reach telephone and vision impairment) Reliability 0.89 No DIF	Validation studies in populations with different levels of disability are needed.
KELS(1988) 2002[49] **2009**[50]	92 Israelian older adults (living in community, living in sheltered houses, living in community but attending day care facilities) 200 community dwelling older adults 65 years or older	17 items	Construct validity Criterion validity (MMSE, FIM, RTI) Construct validity	Construct validity: good discriminative power for the level of dependence of the 3 groups of elderly evaluated. Criterion validity : KELS-MMSE r= -0.757 KELS-FIM r= -0.707 KELS-RTI r= -0.895 KELS-EXIT25 r= 0.705, KELS.-CLOX r= -0.629 KELS-MMSE r= 0.508 KELS-GDS r= 0.318 KELS-mPPT r= 0.472	Predictive validity was not assessed
LL-FDI 2004[53]	101 volunteers aged between 75 and 90 community dwelling adults	32 functional items 16 disability items	Concurrent validity Predictive validity (400-m W and SPPB)	400-mW -function component r= 0.69 p<0.001 400-mW-disability component r=0.26 p> 0.004 SPPB-functional component r= 0.65 p<0.001 SPPB-disability component r=0.16 p>0.004.	Validity over measures of IADL or Cognitive impairment has not yet reported

2.3 New measures of disability based on the world health organization's international classification of functioning disability and health (ICF)

The World Health Organization in 1980 recognized that functioning, disability and health are domains that could describe the general status of a patient in different situations. The International Classification of Functioning Disability and Health (ICF) have been conceptualized to be a framework for the description of health and health related states. The ability of a person to perform some activities depends from his capacity (defined as what the person can do in a standard environment) and from his level of performance (defined as what a person can do in his usual environment).

As already pointed out in the introduction of this thesis, participation in social activities is perceived as a fundamental component of the successful ageing process, moreover it depends on the level of function and on the environment in which the subject lives. Recognizing that the level of disability does not only depend on the level or the severity of the disease but also on the environment, the societal role, the psychological status is a big step forward in the definition of a person's well being by putting new basis in measuring disability and functioning. This more comprehensive model has been defined as the biopsychosocial model and the ICF is based on this model by comprising different perspective of health: biological, individual and social. The opportunity given by ICF is the use of a common language to define disability in every situation considering different and important aspects of the functioning ability. Although the ICF was developed by an International Organization using a consultative process, and is the first useful classification in a variety of situations, it has been pointed out that being so broadly applicable leads to a lower internal consistency. The discriminative power is lower and as per WHO acknowledgement, the difference between activity and participations is not clearly delineated[54]. ICF seems nowadays a good starting point from which to develop measures of disability in different settings and validate new constructs in a variety of populations.

Attempts to develop new concepts to define and measure disability in older adults based on the ICF language are in progress, for example the framework

developed by Freedman comprehends also a domain called “accommodations”. In the case of the National health and Aging Trends Study (NHATS), the domain “accommodations” reflects the behavioral responses to changes in capacity and includes the receipt of help, changes in the environment and other compensatory strategies. This particular domain, as already pointed out in the first chapter, is one of the key characteristics identified by older adults as important to being able to age successfully. Further the ICF classification does not draw a well defined line between ADL and IADL in the domain of activities included, while the NHATS tool defines the ability to carry out self care and domestic life. As with ICF, NHATS recognizes also that the environment may influence the entire disablement process[55]. The NHATS collaborating group is developing a proper measure based on ICF checklist and therefore the validation study is not yet available.

THE INTERATIONAL CLASSIFICATION OF FUNCTIONING DISABILITY AND HEALTH (ICF)

Instead of focusing on subject’s health related disability, the ICF classification focuses on the levels of health by measuring functioning in the society regardless of the reasons for a subjects’ impairment. The ICF classifies functioning and disability associated with health conditions and puts special attention on the impact health has on the global status. Therefore all health related conditions are put at the same level allowing comparisons between them. Disability and functioning are perceived as outcomes of an interaction

between health conditions (diseases, disorders, injuries) and contextual factors. Within the contextual factors it is possible identify external factors (social attitudes, architectural barriers, climate) and personal factors including gender, age, social background, education, profession, character.

The domains in ICF are body functions (including cognitive, psychological), body structure, activities and participation and environmental factors. The list of domains becomes a classification when qualifiers, which record the presence and the severity of a problem, are used. For classification of body functions and structure the qualifier indicates the presence of impairment and its degree. The performance qualifier describes what the subject does in his environment. The capacity qualifier describes the actual subject's ability to execute a task placing an upper limit level to the person's capacity.

In table 3 are summarized scales and indexes based on the ICF checklist, which have been validated in an older adult population and are available in the literature

GERIATRIC CORE SET OF ICF IN AN ACUTE RAHABILITATION SETTING

Identifying the core set of ICF items or domains important to older adults and useful in a geriatric context, is the first step towards a measuring tool based on ICF classification. This was done in a cross sectional study of 150 elderly subjects requiring rehabilitation. According to the results of this study, all 4 ICF components body functions, body structures, activities and participation are

important to the elderly. In the population studied, 82 out of the 125 total activities were chosen from more than 30% of the elderly interviewed. Environmental Factors was the domain with the highest number of items chosen (57%), followed by component of the Body Functions (45%), Activities (35%) and Body Structure (15%). This study, although performed in a post acute setting, represents the first step towards the development of an ICF core set for geriatric patients[56] (see table 4).

The Geriatric Core set has been validated in a post acute rehabilitation cohort of 137 patients. The predictive validation over outcomes of regained ability and dependence after discharge, was analyzed with a non parametric regression method called the Classification and Regression Tree (CART). Particularly, the item d465 ("moving from place to place using equipment") was consistently associated with loss of independence (OR 7.6 95%CI 1.6-35.5). Patients who were reported to have age or disease related involuntary movement disorders, had also a higher risk of being placed in a nursing home (OR 5.6 95%CI 2.6-13.4). This validation study doesn't give enough information on the ICF Geriatric Core set[57].

The World Health Organization Disability Assessment Schedule 2.0 (WHODAS II)

As already conceptualized from the ICF framework, the World Health Organization Disability Assessment Schedule 2.0 captures the individual's level of functioning in six major domains: cognition (understanding, communication), mobility (ability to move and get around), self care (personal hygiene, dressing, eating, living alone), getting along (ability to interact with people), life activities (ability to carry out responsibilities at home, work, school), participation in society (ability to engage in the community, civil, recreation activities). All the items were developed from the ICF comprehensive set. Items were selected after exploring the validity of the ICF items in different cultures, spanning 19 countries. For each of the two studies, participants required for each centre were divided in 4 groups having the same number of subjects. Subjects with different levels and types of disability were selected: subjects with good health status, people with physical disorders, people with mental or emotional disorders, people with problems related to alcohol and drug abuse. At the end of the two phases, 36 items were selected. The WHODAS II was found to have high internal consistency (Chronbach's $\alpha = 0.86$) and a stable factor structure across all countries. The overall intra class correlation (ICC) was very good (0.98). The concurrent validity versus other measures of disability showed, as expected, higher correlation between the FIM and the domains mobility (- 0.78), self-care (- 0.75), household (- 0.60) and participation in society (-0.62). Other scales used for concurrent validity were measures of quality of life and the London Handicap Scale (LHS), which showed good correlation with the

cognition domain (-0.62), household (- 0.64) and participation in society (-0.64)[58].

The WHODAS II, composed of 36 items has been validated in different settings involving patients with different levels of physical and mental disability. A 12 item WHODAS measure has been developed. This new version has the advantage to be administrable in 5 minutes and after a careful review and substitution of less comprehensible items, became easier to answer[59,60]. The 12 item version has been validated in a multicultural setting composed of older adults using the dataset of the 10/66 DRG surveys in 7 developing countries (Cuba, Dominican Republic, Peru, Venezuela, Mexico, China, India). Factor analysis confirmed the presence of one main domain confirming that all the items are at various level measure of disability. The Mokken analysis (which proves the hierarchical sequence of the items) proved the monotonicity of the measure, with item scalability coefficient all higher than 0.4, and most of them higher than 0.5 in all sites. This study confirms that the WHODAS II screening version is unidimensional and hierarchical but more work is needed to prove his reliability, validity in population of older adults with different characteristics[61].

CONCLUSION ON NEW DISABILITY MEASURES BASED ON ICF CLASSIFICATION

The International Classification of Functioning Heath and Disability provides a framework from which to start developing new measures of disability which take

account not only the individual's ability to function in the usual environment, but also in the society and in the community. The quantity and quality of information needed to evaluate the grade of disability of older adults is broader with respect to ADL and IADL scales. The development of a proper scale, able to capture all this information is therefore needed.

Conclusion

This brief review demonstrates the need nowadays for an instrument able to capture all the important aspects of ability and independence as defined by the broader ICF classification, and beside IADL and BADL, also participation in social life as a measure of independence and ability. Scales that have been developed with the purpose of measuring disability are often elaborate and not easy to administer especially considering the needs and capabilities of older adults. Moreover the instruments developed from the ICF framework do not include items aiming to gather information systematically on the cognitive status of the respondent, which has been proved to be often related to disability. The Standard Assessment of Global Activities in the Elderly (SAGE) instrument was developed as a brief and simple scale but comprehensive of all the items recognized to be fundamental to measure ability.

Table 3. ICF BASED MEASURES OF DISABILITY VALIDATED IN OLDER ADULTS

Name, Year, Reference	Population	Items	Validation Method	Validaty-Reliability Results	Methodological consideration
ICF GERIATRIC CORESET 2005[56] 2007[57]	137 older adults from 5 post acute rehabilitation facilities	82 items	Predictive validation with the CART method. Outcome: return to the same habitat enviroment	Predictive value of two particular items moving from place to place using equipment") was consistently associated with loss of independence (OR 7.6 95%CI 1.6-35.5) age or disease related involuntary movement disorders had also a higher risk of being placed in a nursing home (OR 5.6 95%CI 2.6-13.4)	Scale difficult to administer with 82 items. Reliability not reported. Concurrent validity not assessed
WHODAS II 2010[61]	Secondary analysis of the 10/66 DRG surveys in 7 developing countries.	12 items	Factor Analysis Analysis of hierarchical sequence (Mokken method)	One factor in all sites Scalability coefficients > 0.4 in all sites, mostly > 0.5	Validation studies in a generalizable sample of older adults required. Reliability non reported.

Table 4. Categories of the Geriatric ICF core set selected in a acute rehabilitation facility[54]

APPENDIX: *Categories of the geriatric ICF Core Set*	s750 Structure of lower extremity
Body Functions	s760 Structure of trunk
b110 Consciousness functions	s770 Additional musculoskeletal structures related to movement
b114 Orientation functions	s810 Structure of areas of skin
b117 Intellectual functions	*Activities and Participation*
b130 Energy and drive functions	d130 Copying
b134 Sleep functions	d155 Acquiring skills
b140 Attention functions	d177 Making decisions
b144 Memory functions	d230 Carrying out daily routine
b147 Psychomotor functions	d240 Handling stress and other psychological demands
b152 Emotional functions	d310 Communicating with – receiving – spoken messages
b156 Perceptual functions	d315 Communicating with – receiving – nonverbal messages
b167 Mental functions of language	d330 Speaking
b176 Mental function of sequencing complex movements	d335 Producing nonverbal messages
b180 Experience of self and time functions	d360 Using communication devices and techniques
b210 Seeing functions	d410 Changing basic body position
b215 Function of structures adjoining the eye	d415 Maintaining a body position
b230 Hearing functions	d420 Transferring oneself
b240 Sensations associated with hearing and vestibular function	d440 Fine hand use (picking up, grasping)
b260 Proprioceptive function	d445 Hand and arm use
b265 Touch function	d450 Walking
b270 Sensory functions related to temperature and other stimuli	d460 Moving around in different locations
b280 Sensation of pain	d465 Moving around using equipment
b320 Articulation functions	d510 Washing oneself
b410 Heart functions	d520 Caring for body parts
b415 Blood vessel functions	d530 Toileting
b420 Blood pressure functions	d540 Dressing
b430 Haematological system functions	d550 Eating
b435 Immunological system functions	d560 Drinking
b440 Respiration functions	d570 Looking after one's health
b450 Additional respiratory functions	d760 Family relationships
b455 Exercise tolerance functions	d770 Intimate relationships
60 Sensations associated with cardiovascular and respiratory functions	d860 Basic economic transactions
b510 Ingestion functions	d930 Religion and spirituality
b525 Defecation functions	d940 Human rights

b530 Weight maintenance functions	*Environmental Factors*
b535 Sensations associated with the digestive system	e110 Products or substances for personal consumption
b540 General metabolic functions	e115 Products and technology for personal use in daily living
b545 Water, mineral and electrolyte balance functions	20 Products and technology for personal indoor and outdoor mobility and transportation
b620 Urination functions	e125 Products and technology for communication
b630 Sensations associated with urinary functions	e140 Products and technology for culture, recreation and sport
b710 Mobility of joint functions	45 Products and technology for the practice of religion or spirituality
b715 Stability of joint functions	50 Design, construction and building products and technology of buildings for public use
b730 Muscle power functions	40 Light
b735 Muscle tone functions	45 Time-related changes
b755 Involuntary movement reaction functions	50 Sound
b760 Control of voluntary movement functions	10 Immediate family
b765 Involuntary movement functions	15 Extended family
b770 Gait pattern functions	20 Friends
b780 Sensations related to muscles and movement functions	25 Acquaintances, peers, colleagues, neighbours and community members
b810 Protective functions of the skin	30 People in position of authority
b820 Repair functions of the skin	55 Health professionals
b840 Sensation related to the skin	60 Health related professionals
Body Structures	10 Individual attitudes of immediate family members
s110 Structure of brain	15 Individual attitudes of extended family members
s120 Spinal cord and related structures	20 Individual attitudes of friends
s320 Structure of mouth	25 Individual attitudes of acquaintances, peers, colleagues, neighbours and community members
s410 Structure of cardiovascular system	30 Individual attitudes of people in positions of authority
s430 Structure of respiratory system	50 Individual attitudes of health professionals
s610 Structure of urinary system	e455 Individual attitudes of other professionals
s620 Structure of pelvic floor	e460 Societal attitudes
s710 Structure of head and neck region	e465 Social norms, practices and ideologies
s720 Structure of shoulder region	e570 Social security, services, systems and policies
s740 Structure of pelvic region	e580 Health services, systems and policies

Chapter 3 Development and characteristics of the Standard Assessment of Global activities in the Elderly (SAGE) scale

Introduction

The chapter describes the SAGE measurement, by briefly reporting the process through which items were generated and selected. Items in SAGE were subdivided in 4 principal domains: cognition, instrumental activities of daily living (IADL), basic activities of daily living (BADL) and participation. Responses were given by reporting how much difficulty interviewed patients had in the previous month in performing the activities included in the scale.

3.1 Items generation and selection

In the past disability was always measured as a result of level of illness or physical capacity. The new definition of disability by the International Classification of Function Health and Disability gives new input in recognizing what is being important to elderly and it constitutes a framework from which new measures of disability could originate. Scales which are already developed are often elaborate and missing some important aspects such assessing memory (which is the first manifestation of cognitive impairment). After an exhaustive review of the existing scales and literature experts in Theory of Measurement and Elderly Care, including Dr Martin O'Donnell, Dr Mary Law, Jackie Bosch and other colleagues, decided to develop a new instrument to measure disability as a whole construct in the elderly able to capture the switch from ability and disability:

the **Standard Assessment of Global Activities in the Elderly**. The scale is composed of different domains, all considered important and relevant for the elderly. The items are also defined to be (according to expert opinion) “stand alone” items. Components of the Barthel Index (BI) for BADL based on validated 5-item subscale[62], and the Lawton & Brody Scale[63] for IADL were selected. Both the Lawton & Brody and Barthel scales have been validated in a variety of clinical settings and numerous countries. The Barthel scale measures 5 basic ADLs, and has been used widely for numerous clinical conditions and outcome measures in randomized controlled studies on stroke patients. The Lawton & Brody Scale measures 8 domains of IADL, which are used commonly in geriatric research. The presence of ADL and BADL items ensure that SAGE will be broadly applicable to older adults with different levels of disability/ability. Final item selection was based on discussions with a panel of experts in the field.

3.2 Principal Domains

Cognition

The first manifestation of cognitive impairment for vascular cause is the loss of function. Unfortunately, this loss is not captured by validated scales (such as the Mini Mental State Examination)[64]. In the SAGE scale, questions which can assess an early loss of cognition are included.

Participation

According to the World Health Organization (WHO) participation restriction is the direct societal consequence of health issues. Different tasks are considered in the SAGE such as: moving around the neighborhood independently and taking part in social activities.

Instrumental Activities of Daily Living (IADL)

Functions in instrumental activities are lost before basic ADL (bathing, eating, using the toilet), consequently assessment through the SAGE scale may identify incipient decline in an older adult who might appear capable and healthy.

Basic Activities of Daily Living (BADL)

Different tasks which are important for the individual independence are bathing, walking, using stairs and eating with or without help from another person. The order of the items in the SAGE is not random as it ideally reflects the hierarchy of comparative difficulties and measures the participation and the ability to perform ADL across a broad spectrum of everyday life. The ordering of items reflects the expected ordered loss of function (participation in community activities, IADL, BADL). The scale is also developed to be feasible and generalizable cross-culturally in men and women. In the scale, patients are asked to indicate how much difficulty they had in the last 30 days in performing different activities. In the case of IADL and BADL, it is also required to indicate if the help of another person (for example to manage and take medications) or a tool (for example to walk or climb the stairs) was needed. The SAGE scale has been developed to be

useful in assessing the grade of abilities over a broad spectrum of activities in community dwelling adults and older patients with some degree of disability. Therefore a proper process of validation of the assessment in this type of population is needed.

3.3 Responses

Older adults are asked to directly estimate the difficulty they had in the past month with the activities listed in the questionnaire by choosing between none, mild, moderate, severe or it is possible indicate whether the activity is not performed at all. Each level of difficulty is defined on the basis of the intensity of the difficulty or the frequency in which the individuals finds difficult to perform a task. It is known that recalling behavior or difficulties leads often to an underestimation or overestimation of the phenomenon based on the memory of the respondent. This is usually true when the behavior occurred more than a month before, or with difficulties that fluctuate over time[65]. In the case of the SAGE the tasks for which difficulties are reported are everyday activities and therefore these should not constitute a possible source of recalling bias.

Mild difficulty is defined as minimal or occasional that does not affect the ability to perform the particular activity. While some or regular difficulty that does effect the ability to perform the task but the subject is still able to complete the task describes a **Moderate** difficulty. Finally a **Severe** difficulty is defined as an extreme or constant difficulty in performing the task. The severe difficulty answer

could also be endorsed in case the task is performed form someone else because of its difficulty.

3.4 Biases in responses

There are possible sources of bias in responses. The respondent answers could be influenced for different reasons. In this case, assessed older adults may report a frequency of performed activities that it's desirable but not necessary true. This phenomenon is called "social desirability" and the reasons for reporting a biased answer could be different. One of the main reasons could be the willingness to hide the real ability status or the wish to please the assessor. On the contrary some older adults may want to exaggerate their real condition. Self reported outcome could also be influenced by mild cognitive impairment of the respondent and in this case the evaluation of the respondent with the MoCA test will be very useful to detect this condition.

Conclusions

The development of SAGE followed different considerations. Firstly the definition of disability is nowadays changed, independence is measured not only on the basis of activities of daily living, but also on the possibility for the older subject to take part in social and community life. Secondly a brief and simple tool, to be administered to a homogeneous population including community dwelling or institutionalized older adults was needed. The SAGE item selection and development was completed by the panel of experts in the geriatric field of

McMaster University. In this thesis I will address the methods needed to study the SAGE's validity and reliability.

Figure1.The SAGE SCALE

SAGE Validation Standard Assessment of Global Activities in the Elderly (SAGE)

DataFax #078 Plate #024 Visit #001

Participant ID# ☐☐☐☐ ☐☐☐☐ *Centre No.* *Participant No.*

Initials ☐☐☐ *F M L*

Date completed: ☐☐☐☐ ☐☐ ☐☐ *year month day*

Completed by: ☐ Participant
☐ Other person
Relationship to participant: _____

Activity

Over the past month, how much difficulty have you had with the following:

	Level of Difficulty *(please check only one response)*				
	None	Mild	Moderate	Severe/stopped due to difficulty	Never performed
1. Keeping your attention or 'train of thought' during a conversation?	☐	☐	☐	☐	☐
2. Remembering things that happened a few days before? (e.g. conversation, people visiting)	☐	☐	☐	☐	☐
3. Playing a game or reading a book that requires concentration? (e.g. of games: crosswords, checkers, chess)	☐	☐	☐	☐	☐
4. Ability to switch between things that are happening at the same time? (e.g. making tea and talking to someone)	☐	☐	☐	☐	☐
5. Finding your way around a new building? (e.g. hospital/clinic)	☐	☐	☐	☐	☐
6. Organizing a trip or social activities? (e.g. vacation or family occasion) (score the activity that the person finds to be the more difficult of the two)	☐	☐	☐	☐	☐
7. Doing your own finances or shopping? (score the activity that the person finds to be the more difficult of the two)	☐	☐	☐	☐	☐
8. Organizing and taking your medications?	☐	☐	☐	☐	☐
9. Preparing a meal and/or doing laundry? (score the activity that the person finds to be the more difficult of the two)	☐	☐	☐	☐	☐
10. a) Driving If the patient does not drive, please answer question b)	☐	☐	☐	☐	☐
→ b) Using public transportation?	☐	☐	☐	☐	☐

Participant ID# ☐☐☐☐ ☐☐☐☐ **Initials** ☐☐☐
Centre No. *Participant No.* *F M L*

Over the past month, how much difficulty have you had with the following:

Level of Difficulty
(please check only one response)

	None	Mild	Moderate	Severe	Do not perform this activity
11. Using stairs? (one flight)	☐	☐	☐	☐	☐
12. Walking? (approx. 10m or 32ft or 14 steps)	☐	☐	☐	☐	☐
13. Dressing?	☐	☐	☐	☐	☐
14. Transfer from bed to chair?	☐	☐	☐	☐	☐
15. Bathing or toileting? (score the activity that the person finds to be the more difficult of the two)	☐	☐	☐	☐	☐

→ Do you require help from: ☐ Walking stick/rollator ☐ Another person ☐ Elevator/lift ☐ Other: _____

→ Do you require help from: ☐ Walking stick/rollator ☐ Another person ☐ Elevator/lift ☐ Other: _____

→ Do you require help from another person? ☐ Yes ☐ No

Score ☐☐

Time to administer ☐☐ **mins**

If the participant reports difficulty with any of the items in questions 6-15, please ask:

16. Have any of the following problems limited your ability to perform activities? *(check all that apply)*

☐ Memory problems	☐ Physical injury (e.g. fracture)	☐ Loss of vision	☐ Not sure
☐ Arthritis	☐ Stroke or TIA	☐ Unsteadiness	
☐ Shortness of breath	☐ Chronic pain	☐ Other: _____	
☐ Chest pain	☐ Heart failure	☐ None of the above	

Did you find the questions easy to understand and respond to? ☐ Yes ☐ No

Was there anything that you think was missing from the questions? ☐ No ☐ Yes _____

Would you have changed the order, wording or anything about the questions? ☐ No ☐ Yes _____

Person Completing Report: _____ Date: ☐☐☐☐ ☐☐ ☐☐
Print last name/first initial ***Signature*** *year* *month* *day*

Draft 12 - Mar. 01, 2010

Introduction

In the developing process of a new tool or scale it is important assess the reproducibility of the results (reliability), but it is also very important to assure that the conclusions that can be drawn from the evaluation of a subject with the new scale are accurate. Moreover, a tool has to be reliable in describing and discriminating between patients with different degrees of a disease or disability. The purpose of the SAGE scale is to measure the degree of disability of older subjects through the evaluation of different aspects of their independence. In this chapter history and development of “validity” will be discussed.

4.1 What is validity ?

In measurement theory an instrument is valid if the conclusion that can be drawn from the administration of the tool reflects the real status or condition of the respondents. As outlined in the previous chapter, the SAGE contains items which aim to measure the degree of cognitive function, the ability of the subject in participating in social and community activities and the independence in performing basic and instrumental activities of daily living. It is fundamental to assess the validity of SAGE mainly for two reasons. Firstly, it is important to make sure that the SAGE is really measuring disability in older people, in other words, we need to be sure that the items included in SAGE are comprehensive and collect as much information as possible on different aspects which are important to older subjects to be independent. Secondly, physicians and health

care professionals need to be sure that the conclusions drawn on the basis of the information collected through the SAGE questionnaire, really reflects an older adult's level of disability.

4.2 Why it is important to assess SAGE validity

Disability as measured by the SAGE scale is the result of the interplay of different health and cognitive losses but it is also influenced by personal, environmental and social factors. This can lead to participation restriction and activity limitations. As pointed out by WHO: "*Disability is a complex phenomenon, reflecting an interaction between features of a person's body and features of the society in which he or she lives*"[66].

In this case therefore, it is not a physical quantity such as body temperature or blood pressure which is by definition readily observable and can be directly measured, but a construct to be assessed. A construct is a property, an underlying characteristic that the scale is purported to measure but is not directly measurable. Measures of cognitive function, basic and instrumental activity of daily living and participation are concepts that may vary from patient to patient and from the way they are measured. Validity assessment is important to assure, for example, that patients who score lower in the SAGE scale are really older adults with little or no disability at all. In the case of SAGE, validity is also required because the construct of interest is not directly measurable, and results from collecting information on the physical, cognitive and social status of older

adults. SAGE will be a useful tool if it is demonstrated that the conclusion drawn on the disability degree of a subject, on the basis of the scores obtained with SAGE, accurately reflects the subject's disability status. In other words, it's not the scale to be validated but the score obtained through it. Indeed the meaning of the score varies on the basis of the population and the setting from which it has been derived, validity measure is directly related to the same parameters[67].

4.3 Types of Validation studies

The three "C"s

In the 1950's scientific discussions (mainly based on psychology) focused around validity of a test that would lead to only verify the accuracy of the estimate obtained with the test. This model suggested the use of a reference measure to compare the new tool to. The first way through which to assess validity was called "criterion validity" because of the presence of a "criterion" that was considered the reference [68].

CRITERION VALIDITY

Traditionally the "criterion validation" is how the score obtained with a new tool correlates with the score obtained in a comparable population and setting with another test, designed to measure the same property or task. The reference test is defined as the "gold standard". Criterion validity is tested using Pearson's correlation which is a measure of correlation of two continuous variables. The

higher the Pearson's correlation, the better the two variables are correlated with each other and is, numerically, a reflection of the same underlying property. Criterion validity can be tested:

a) At the same time as the gold standard (concurrent -criterion validity)

b) The new test can be correlated with an indicator or a property that will be assessed in the future (predictive-criterion validity)[68,69]

Issues related to the criterion validation can arise in fields in which no gold standard is available. In some cases, even though reference tests exist, doubts about the population characteristics or the setting in which those criteria are validated could present some concerns[70].

CONTENT VALIDITY

Content validity has been defined as the extent to which the new scale represents or covers all the different aspects of the construct to be measured. For example, if the measure is incomplete and does not cover one important aspect of the outcome, then the inferences that can be drawn are most likely to be wrong (or incomplete). Therefore the higher the content validity, the broader the inferences can be draw on the subject in a variety of different situations. Generally, content validity has been based on a review of the measure's tests by experts in the field, but also other methods are used to assess content validity such as by calculating internal consistency or through factor analysis (the last two

approaches will be outlined in chapter 6). As already stated above, the construct of interest could be characterized by aspects that might be different from patient to patient. Moreover the judgment on the appropriateness of the scale could be subjective and often done by the same researchers involved in the development of the scale.

CONSTRUCT VALIDITY

As reported by Cronbach and Meehl back in 1955[69], Construct validity is studied when the tester has no prior definite gold standard or criterion, so indirect measures to individuate the underlying construct are needed. Moreover construct validation is involved whenever a test is to be interpreted as a measure of some attribute which is not directly measurable.

While content and criterion validity can be assessed with one or two experiments, Construct validity is believed to be an ongoing process of learning more about the construct, by hypothesizing and testing new theories about it[70]. To better define this, using Chronbach and Meehl's words, the process of validation of a new construct means elaborating the "nomological net" in which it occurs. A Nomological net is the set of laws that describes the theory to be proven. These laws can correlate observable and measurable quantities, the former to theoretical constructs, or finally relate theoretical constructs to each other. When the net is enriched with new construct or new relations to theory, it is necessary

that these additions generate laws that are confirmed by observations or that reduce the number of laws needed to predict the same observations[71].

Construct validity can be measured simply by assessing two different groups (one with the trait intended to be measured and one without), with the scale to be validated. It should be expected that the results would be very different. This method is called “construct validation by extreme groups” and it becomes very useful in the process of developing a new tool. This method presents some difficulties: if the aim of the investigator is developing a new and better tool, it can be difficult to find a way to select the extreme groups. Usually extreme groups are selected by using the best available tool or expert judgment or a scale tapping most of the features the new tool is intended to cover. Therefore if the new tool allows better interpretations or explains more findings, the gold standard can be replaced[67].

Assessing the close relationship between the new tool and variables or measures related to the same construct, is also the purpose of “Convergent Construct Validation”. At the same time the correlation between the two scales of measurement can’t be too high, otherwise there is no difference in using one or the other. On the contrary, it would be expected that the new scale would not correlate with dissimilar or unrelated measures, which is called “Discriminant Construct validation”.

Nowadays, construct validity is generally considered as a unifying form of validity, comprising both content and criterion validity which traditionally were considered as separate and distinct forms.

Messick was the first in 1995 declaring that:

" *Construct validity is based on an integration of any evidence that bears on the interpretation or meaning of the test scores –including content and criterion-related evidence- which are thus subsumed as part of construct validation*"[71].

According to Messick, the so called "Trinitarian view" discussed above should not exist and all types of validity testing (not different varieties of validity) are all aspects of Construct validity. There are no more "content validity", "criterion validity" but "content validation" and "criterion validation" processes to study validity as a singular construct.

In the same article Messick also indentified six contributors to construct validity:

a) Content aspects

Content aspects are important to identify the pool of aspects that have to be revealed or tapped by the measurement (Relevance). Moreover the assessment should include items that are representative/characteristic of the underlying construct (Representativeness). Usually both content relevance and representativeness are addressed by expert judgment.

b) Substantive aspects

Substantive aspects represent a step forward with respect to content aspects of construct validity. Messick highlights the need for the scientific proof that the theoretical processes intended to be measured, are engaged by respondents. These evidences could derive, for example, from consistencies in response times.

c) Structural aspects

Structural aspects are related to the scoring model used in the tool. The scoring model should be derived from what is known about the trait intended to be measured. Therefore the construct object of measurement and the structural net underlying its manifestation have to drive not only the content selection but also the scoring model.

d) Generalizability aspects

Generalizability aspects appraise the extent to which scores obtained with the new tool are broadly generalizable across population groups and settings. In other words score inferences do not have to be limited only on the population and in the setting of assessment.

e) External Aspects

The External Aspects refer to the correlation or non correlation with scores obtained from measures of the same construct or measures of different construct respectively. Both “convergent” and “discriminative” aspects are informative and

important. Multi trait multi method matrixes (MTMM) are useful ways to evaluate both convergent and discriminant aspects at once. In MTMM two or more different traits are measured with two methods (for example by rater and by examination). In the matrix there will be correlations calculated for the same trait (homotrait) measured with different methods (heteromethods), different trait (heterotrait) measured with the same method (homomethod) and the different traits (heterotraits) measured with different methods (heteromethods).

Convergent aspects of construct validity will be highlighted from the homotrait-heteromethod: the same trait measured with different methods should give highly related scores. While discriminant aspects of construct validity are shown by low correlation of homomethod-heterotrait.

f) Consequential Aspects

Consequential aspects of construct validity are related to the implications to the score interpretation. For example a score derived from a new test could be supportive in the decision making process or in clinical judgment in the short or long term.

Conclusion

Assessing validity of a new measure is a process that depends from different factors. First of all validation is dependent on the population involved in the study, which inferences can only be translated in the same population. Secondly validity

is usually tested with the support of a gold standard, and sometimes (such as in the SAGE case) no scale has already been validated and can be used as comparator. Finally validity is an ongoing process, there are infinite ways and situations in which validity can be assessed. Particularly in the process of validating the SAGE scale in a population of older adults, content validation, construct validation and criterion validation will be assessed. The methods through which assess this aspects of validity will be described in chapter 6.

Chapter 5 Design of the SAGE Validation study

Introduction

The aim of this preliminary study is to test the validity of the SAGE as a measure of disability in the elderly. In this chapter the appropriate study design, the intervention, the outcome of interest, the population selected and the method of sampling will be described.

5.1 selecting a study design

Items in the SAGE have been selected to be applicable to older adults with different characteristics and with different levels of independence. Therefore it is expected that the SAGE will be broadly applicable in community dwelling older adults and those with different degrees of disability.

The design of this preliminary validation study will be cross sectional. The cross sectional design has some advantages and disadvantages. First of all, it allows for very quick evaluation of all the subjects in a study at one point in time. In this population, collecting quickly as much information as possible is important. In older subjects, co-morbidities and disabilities could worsen in a very short period of time, thus changing outcomes of interest. Cross sectional studies are also fast and inexpensive. The disadvantage is that cross sectional designs are useful in describing a population but not in creating inferences on causality of events[72]. Patients will be evaluated at one point in time with the SAGE scale. At the same

time baseline characteristics such as patient's medical information and also family history of cardiovascular risk factor will be collected, but subsequent developments will not be available.

5.2 Method

The method will consist of administering to the selected population the SAGE scale and concurrently four already validated and widely used scales for the assessment of physical disability and cognitive impairment in the elderly. Nowadays a proper tool which has all the four sub-domains included in SAGE and can be used as a gold standard is not available in the literature. Therefore in order to measure concurrent validity of the new tool in the validation phase of the study a measure of cognition (The Montreal Cognitive Assessment, MoCA), a measure of function (The Franchay Activity Index, FAI) a measure of general outcome used principally in post stroke patients (the Modified Rankin Scale, m-Rankin) and a measure of frailty (the CSHA Clinical Frailty Scale) will be administered to all subjects participants.

<u>The Montreal Cognitive Assessment (MoCA)</u>

The Montreal Cognitive Assessment (MoCA) is a cognitive screening test, thought to be sensitive and applicable in cases of Mild Cognitive Impairment (MCI) in patients that usually perform in the normal range of the Mini Mental State Examination.

The MoCA was designed to be rapid (it takes 10 minutes to be administred) and simple. It assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuo-constructional skills, conceptual thinking, calculations and orientations. The total possible score is 30 points, and a score of 26 or above is considered normal. The MoCA scale has been validated in many languages and in populations with different levels of cognitive impairment. The scale was first validated in three groups of patients: patients with Alzheimer's disease (AD), patients meeting criteria for MCI and normal elderly controls (NC) each consisting of about 90 subjects. Internal consistency revealed a Chronbach α of 0.83, test retest reliability was 0.92 ($p < 0.001$). All items were successful in discriminating between the three groups of patients. Compared to MMSE, the MoCA test was much more sensitive in detecting MCI when the cut off score was set at 26 points (90% versus 18% in the MMSE) and in the Alzheimer Disease group MoCA had a sensitivity of 100% compared to MMSE (78%). MoCa is therefore a valid tool to evaluate and discriminate levels of cognition in patients presenting with cognitive complaints[73].

Rasch analysis on a sample of 222 outpatients from a geriatric clinic, most of whom were mildly cognitively impaired or demented, showed the unidimensionality of the construct in measuring cognitive functions. As internal consistency, the authors decided to report the person separation index which provides an estimate of the reliability of the measure taking into account the error in estimating the persons' ability. The person separation index was estimated to

be equal to 0.749. The MoCA test is a reliable instrument to evaluate the cognitive ability in the elderly. Further investigation must be taken to understand which change in the MoCA score is clinically meaningful[74].

Franchay Activity Index

The Franchay Activity Index has been already discussed in the section "background on available scale". In summary the FAI includes 15 items and requires the self reported frequency with which the patients carry on these activities. The activities included are both BADL (such as preparing meals, washing up) and IADL (housework, local shopping, driving a car)[45].

Modified Rankin Scale

The original Rankin scale was developed by Dr John Rankin in years of observation of patients affected by stroke. The scale was composed of 5 hierarchical grades of "functional recovery" ranging from Grade I described as "no significant disability" to Grade V "severe disability"[75]. The Modified Rankin Scale (mRS) was firstly introduced as an outcome measure of the severity of stroke in the UK-TIA trial. The Modified version ranged from 0 which was described as "no symptoms at all" to 5: "severe disability", in the UK-TIA trial strokes scoring from zero to two were counted as non disabling and those scoring from three to five were counted as disabling. The mRS was used to evaluate the disability status of 1717 post stroke patients evaluated at 90 ± 7 days after the event. The mRS

categories significantly correlated to the mean time patients spent at their own homes or at relatives' homes after the hospitalization for stroke[76].

The mRS is the most used scale in stroke clinical trials. Studies conducted on the reliability of this scale demonstrated good test retest reliability, measuring a similar weighted k agreement ranging from 0.96 to 0.99. Inter-rater reliability was also proved to be moderate to nearly perfect among 3 separate studies ranging from a weighted k of 0.71 to a maximum of 0.93[77].

The CHSA Clinical frailty scale

The Clinical Frailty scale has been developed in different stages of the Canadian Health and Ageing study. This tool measures the overall clinical frailty basing the classification of the frailty state on fitness and function. Classifications are ranging from 1 (robust health) to 7 (complete functional dependence on others). The Clinical Frailty Scale has been validated in a prospective cohort study and it showed a good predictive value against death and need for recovery in an institution. Indeed for each category increment in the Clinical Frailty Scale of 1 a 21.2% increased risk of death within 70 months and a 23.9% increased risk of entering in a skilled facility were observed. [78]

5.3 Sample size

By hypothesizing the desired concurrent criterion validation correlation coefficient, it is possible calculate the sample size needed of each group of older

adults involved in the validation study. A Pearson's correlation coefficient of 0.7 and more is usually considered an indicator of strong correlation. Since a situation of no correlation at all (r= 0) is not real in this context, the null hypothesis (H_0) will be stated as the presence of a weak correlation (r= 0.5) existing between SAGE and the 3 scales used for criterion validation. Considering a power of 80% and a two sided α= 0.05 the sample of older adults needed to be screened in each facility will be 80 (calculation obtained with nQuery Advisor version 6.01).

5.4 Population

The collaborative group was interested to test the tool in group of patients with different clinical status and consequent cognitive and functional ability. The aim was to evaluate the ability of SAGE to detect and respond differently on the basis of the degree of functionality and dependence. Therefore three facilities of the Hamilton area were identified for their possibility of being in contact and assess community dwelling subjects, patients recovering from a stroke and patients permanently hosted in a nursing home.

-HOPE 3 STUDY

The Heart Outcomes Prevention Evaluation (HOPE-3) study is an ongoing large simple, double blind trial with a factorial design on the effect of rosuvastatin (10 mg), Candesartan-Hydrochlorotiazide (16-25 mg) versus placebo in reducing major cardiovascular events over an average follow up of 5 years. The multicentre international trial is coordinated at the Population Health Research

Institute and includes men (aged 55 years or older) and woman (aged 65 years or older) cognitively healthy with another risk factor but not clear cardiovascular disease. The Hamilton site of the Hope 3 study provides a perfect chance to assess healthy community dwelling subjects with no apparent cognitive and functional impairment. Baseline characteristics were collected as part of the HOPE 3 case report form.

-Intensive Care Unit (ICU) Hamilton General Hospital

Post stroke patients can experience different levels of cognitive impairment and a decreased independence to perform activities that prior to the cerebral infarct would be considered usual normal activities. Therefore patients recovering after a stroke episode at the Neuroscience and Trauma ICU of the General Hospital constitute a suitable sample to test the property of SAGE in a population with probable or possible functional impairment. Baseline characteristics will be also collected. Where a direct interview of the patients will not be possible, it is planned to interview a parent or a caregiver (proxy) informed on the status of the patient. In a recent review it has been reported that, although most studies showed that the proxy overestimated impairments compared to patients self reports, in post stroke patients the reliability of the proxy's information depended on the nature and the objective of the questions. The reliability of the proxy respondent for validated scales of ADL ranged between 0.61 and 0.91[79].

-ST PETER'S HOSPITAL

St Peter's Hospital offers inpatient and outpatients care for seniors with acute and chronic illness. In particular the field of research focuses on dementia care, palliative care and rehabilitation in the ageing population. Patients recruited from this facility are considered severely impaired either cognitively and functionally. In this case the use of caregiver information is practically constant, generally patients hosted in a nursing home are enough impaired to require constant help and therefore are considered extremely dependent.

5.5 Sampling process

In every kind of study it would be ideal to test the entire population. Practically this is not possible, because time and resources consuming. Sampling is the process of selecting individuals from a defined population with the purpose of studying the sample and be able to make inferences on the whole population.

Subjects to be included in the validity study of SAGE will be conveniently sampled. Convenience sampling is characterized by a selection of subjects that are closest at hand, readily available and convenient. Usually with convenience sample participants are included in the study consecutively, which means that in the SAGE validation study 80 consecutive patients will be selected and interviewed in each facility. This approach has the advantage of being fast and inexpensive. In some studies convenience sampling has also the advantage of minimizing voluntarism: subjects that want to be selected for the study are usually healthier than the general population and this may distort the results. However

researchers can't reliably draw inferences on the general population by generating hypothesis on a convenient sample since the sample is not representative of the entire population.

Conclusion

In the process of this validation study, to be included will be subjects with different degrees of disability. The aim is allow for inferences on the ability of the new instrument in describing these different characteristics. The study design will be cross sectional, permitting a very quick and complete collection of the data needed. The calculated number of subjects needed to be involved in the study is in total 240. Participants will be conveniently sampled.

Chapter 6 Assessing SAGE Validity

Introduction

In this chapter methods used to assess validity of the SAGE scale will be evaluated. The process of validation will include: measure of reliability, content validation, construct validation, criterion validation, testing the real hierarchical sequence of the items and testing sensitivity and specificity.

Reliability is a measure of consistency of the results obtained with the new scale over time (test retest reliability) and by different raters (inter-rater reliability). Validity, as already discussed previously, is the ability of the new instrument to measure what is really intended to measure. To test the hierarchical sequence of the items it is necessary to prove that endorsing one item on a scale, increases the probability of endorsing less difficult items. Sensitivity is the ability of a test to detect a disease, while by specificity is intended the ability of a test to be specific for one disorder only.

6.1 Primary outcome of this study

The primary outcome of this validation process is proof that SAGE is a valid instrument to detect disability (both cognitive and functional) in a heterogeneous group of older adults with different degrees of illness. As proposed by the WHO,

disability is a broad concept that involves physical, cognitive and social status of a subject. SAGE is the first brief scale developed with the aim of detecting changes in these important aspects of independence. The scale has also been developed to be easy to administer, especially considering the setting in which it will be implemented.

The hypothesis will be that scores obtained with SAGE will strongly correlate with the scores obtained with all the validated and broadly used scales that will be considered as gold standards, that the items are hierarchically ordered from the most to the least difficult and that SAGE is sensitive and specific in detecting disability.

6.2 Data Collection

Data will be collected at one point in time. Patients or caregivers will answer a standardized questionnaire collecting clinical history and demographics (age, sex and ethnicity) information.

Clinical History

Information collected on clinical history will include:

- Presence of Hypertension defined as history of hypertension, measured Systolic Blood pressure higher than 140 mmHg and a Diastolic Blood pressure higher than 90 mmHg or only blood pressure lowering agents, and number of years since the diagnosis

- Presence of history of high cholesterol level, defined as a total cholesterol level higher than 200 mg/dL or on cholesterol lowering medications and the number of years since the diagnosis
- Presence of history of diabetes, defined as fasting plasma glucose levels higher than 126 mg/dL or on treatment for diabetes and the number of years since the diagnosis
- Previous history of hemorrhagic or ischemic stroke and the median number of years since the diagnosis
- Previous history of Transient Ischemic Attack (TIA) and the number of years since the diagnosis of the disease.
- Previous history of Atrial Fibrillation (AF) and number of years since the diagnosis of the disease
- Previous history of Acute Myocardial Infarction (AMI) and the number of years since the diagnosis.

Current medications

It is expected that the population involved in the SAGE validation will likely suffer from more than one chronic disease and will, therefore, be under treatment with different medications. Information on the medical treatments used at the moment of the administration of the instrument will also be collected.

Information on medical treatments will include:

- Oral Hypoglycaemic drugs

- Statins
- Oral Anticoagulants
- Insulin
- Antiplatelets
- Blood Pressure lowering medications
- Antidepressant.

6.3 Scoring System

SAGE

SAGE comprises 15 items, each describing an activity for which the respondent has to indicate how much difficulty he/she has encountered in performing this activity in the past month. The scoring system for SAGE has been decided to be incremental with the increase in the reported difficulty encountered in performing the activities described. Therefore we will assign:

- 0 points if the participants endorse the “None/never performed” response
- 1 point to the “Mild” response
- 2 points to the “Moderate” response
- 3 points to the “Severe” response.

One additional point will be assigned when at question 11, 12 and 15 the subject declares the need for help from another person or a tool to walk, jump the stairs or to bath. The scores will range from 0, describing a very independent

participant over a broad spectrum of activities, to 48 describing a very dependent subject.

Montreal Cognitive Assessment (MoCA)

The Montreal Cognitive Assessment is a widely used scale for the evaluation of dementia and mild cognitive impairment. It Comprises 30 items, assessing different cognitive domains[80] and it takes approximately 30 minutes to be administered. For each task correctly completed, one point is assigned. Therefore the range of scores could be between 0 (for a totally cognitive impaired subject) to 30 describing a cognitively healthy participant. Studies on the validity of MoCA showed that a score under 26 could indicates a mild cognitive impairment, while a score over 26 is considered normal[81].

Franchay Activity Index (FAI)

The Franchay Activity Index has been developed with the scope to measure how often in a period of time, varying between three or six months, the respondent has undertaken various activities (domestic skills, activities to be performed during work or leisure time or outdoor activities). Only in the case of gardening and household activities the respondent has to indicate the intensiveness of the work. There are 15 activities and responses will be scored as reported in the original paper of the FAI scale:

- "None/Never" will be scored with 0

- “Less than once a week”, “1-2 times in 3 months”, “1-2 times in 6 months”, “light gardening or housework”, “1 in 6 months” and “Up to 10hrs per week” will be scored with 1
- “1-2 times/wk”, “3-12 times in 3 months”, “moderate gardening or housework”, “less than 1 every 2 weeks”, and “10-30 hours per week” will be scored with 2
- “Most days”, “at least weekly”, “at least every two weeks”, “all necessary”, “more than one every two wks”, and “over 30 hours per week” will be scored with 3

Scores obtained with FAI can range from 0 to 45, where a score of 0 describes a participant who does not undertake any of the activities listed and 45 describes a very active subject.

Modified Rankin (mRS)

The Rankin scale was developed to measure the change in function of patients rehabilitating from an episode of stroke. Items in the modified Rankin scale are simply description of the level of independence of the subject. Items are scored as follows:

- “No symptoms at all” will be scored with 0
- “No significant disability despite symptoms; able to carry out all usual duties and activities” will be scored with 1

- “Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance” will be scored with 2
- “Moderate disability; requiring some help but able to walk without assistance” will be scored with 3
- “Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance” will be scored with 4
- “Severe disability; bedridden, incontinent and requiring constant nursing care and attention” will be scored with 5.

The mRS comprises also a last item “Dead” which in our case will not be taken into account. The possible scores will range from 0 indicating a subject recovering perfectly to 5 indicating a subject completely dependent.

The CSHA Clinical Frailty Scale

The CSHA Clinical Frailty scale has been described in the previous chapter. Briefly it measures the overall clinical frailty basing the classification of the frailty state on fitness and function. Classifications are ranging from 1 (robust health) to 7 (complete functional dependence on others).

6.4 Statistical considerations

As previously discussed the validation process of the SAGE will follow a cross sectional design evaluation of a group of patients with different degree of

disability. This is also a preliminary phase, the results of which could help design a validity study with the intention to further assess predictive validation in a longitudinal cohort or subjects.

Data presentation

Descriptive data on baseline characteristics of subjects involved in SAGE will be reported overall and by sex for all the important variables. Scores describing the functional and cognitive status of the subjects involved in the validity assessment need to be further investigated.

Normally distributed data

The type of data that will be collected with SAGE are ordinal. Most of the statistical tests that will be used to assess validity of SAGE are functioning under the assumption that the scores' distribution follows a normal trend, called the Gaussian distribution characterized from a bell shape. The normal distribution has been observed in numerous parameters measured: all the points are symmetrically distributed around their mean. The central limit theorem states also that if we draw a sample from a not normally distributed population, the distribution of the means of the sample will be normally distributed as long as the sample is large enough. The rule of thumb is that a sample size over 30 is usually good enough to lead into a normally distributed set of parameters. More importantly, with normally distributed data, the variance remains equal when the mean changes[82]. Therefore the data gathered with SAGE will be tested for

normal distribution by performing a simple descriptive analysis of the distribution of the scores in the entire population and by centre.

As we will see measuring variance is the primary scope of assessing reliability and validity of a new instrument. Variance (σ^2) of a measure is obtained by summing the squares of the differences between the observed individual values and their mean divided by the number of observations.

- Variance $\sigma^2 = \frac{\sum (X - \mu)2}{N}$

In order to assure that scores obtained with SAGE, MoCA, FAI and mRankin follow the trend of a normal distribution, data will be presented as median and interquartile ranges overall and by centre of recruitment. Scores will also be reported stratified by sex and by age intervals (≤55, 56-65, 66-75, 76-85, > 85). It will be expected that the median value score for each scale considered, will change on the basis of the characteristics of the group and the age interval considered.

Variance testing as a measure of validity and reliability

Variance observed in a measure, is the result of different type of variances all playing a role in the final value. Generally variance observed is the sum of the real variance and some error.

$\sigma^2_{observed} = \sigma^2_{true} + \sigma^2_{error}$

In the development and testing of a new scale, particular attention must be paid to the assessment of the Reliability and Validity of the scale, as measures of errors of the observed scores.

As already stated at the start of this chapter, the term reliability refers to the consistency of the scores obtained with the new measure over time (test-retest reliability) or by different raters (inter rater reliability), while validity refers to the extent to which the new instrument is measuring the trait is supposed to measure. Moreover Validity is also directly related to Reliability, the higher the reliability of a new instrument the higher the validity.

In a construct of interest the variance observed $\sigma^2_{observed}$ is equal to the sum of the true variance of the construct of interest plus some systematic error (which occurs systematically due for example to the rater or to the particular occasion in which the observations were done) and some random error (RE)that is not predictable.

A systematic error (SE), given for example from the interplay of Observer error and Item error, is included in the numerator of the Reliability coefficients, because it explains the dependability of inferences from the Observer and the Items of the test. The magnitude of systematic errors tends to increase Reliability.

$$\text{Reliability} = \frac{\sigma^2 true + \sigma^2 SE}{\sigma^2 observed}$$

In contrast, in Validity assessment the systematic error is part of the observed variances and its magnitude tend to decrease validity. Indeed as Validity is the extent to which our observed score is really measuring the trait of interest and it is not resulting only from errors then it would have been proved that SE and RE are only small parts of what observed and the ratio between the error of the construct and what is observed is hopefully near one[83-84].

Therefore:

$$\text{Validity} = \frac{\sigma^2_{true}}{\sigma^2_{observed}}$$

6.5 Assessing internal consistency of SAGE

Prior to validity assessment it is necessary to assess the internal consistency of the new scale. Internal consistency is a measure of homogeneity of the items constituting the scale. It estimates reliability based on the average correlation among items within a test. In general in a new scale, items should moderately correlate with each other and each one of them should correlate with the total scale score. These two aspects are components of internal consistency. Internal consistency is measured with Cronbach's Alpha (α) which is defined as follows:

$$\alpha = \frac{K}{K-1}\left(1 - \sum_{i=1}^{k}\frac{\sigma i\ 2}{\sigma t\ 2}\right)$$

K is the number of items in a scale, σ_i the variance of one item and σ_t the variance of the total score. The coefficient α is dependent not only from the correlation between items but also from the number of items, the higher the number of items the higher the internal consistency. If α is very low the test is either too short or the items are not correlated with each other.

The acceptable value for internal consistency varies from 0.70 to 0.90. As reported recently by Ponterotto and Ruckdeschel, resulting α coefficient should be considered acceptable also on the basis of the sample size involved and the number of items composing the instrument[85] (see table 5). According to the reported values, considering the number of items in SAGE and the estimated sample size, acceptable SAGE internal consistency should range from 0.85 to 0.90.

Although the purpose of SAGE is assess disability, by measuring different aspects of this trait (such as cognition, physical impairment and lack of participation), it should be considered an analysis of internal consistency also by domain to evaluate the grade of correlation of each item within the different domains of the scale. Internal consistency will be reported on the overall population and by centre. In the latter case the expected α will be lower than in the overall population, ranging from 0.75 to 0.85

Table 5 Desirable values of α on the basis of number of items and respondents involved

Items per scale or subscale	Rating	N< 100	100-300	> 300
< 6	Excellent	0.75	0.80	0.85
	Good	0.70	0.75	0.80
	Moderate	0.65	0.70	0.75
	Fair	0.60	0.65	0.70
7-11	Excellent	0.80	0.85	0.90
	Good	0.75	0.80	0.85
	Moderate	0.70	0.75	0.80
	Fair	0.65	0.70	0.75
> **11**	Excellent	0.85	0.90	0.90
	Good	0.80	0.85	
	Moderate	0.75	0.80	0.85
	Fair	0.70	0.75	0.80

6.6 Validity assessment

CONTENT VALIDITY

Content validity will be assessed by expert opinion and factor analysis. Experts in the field of geriatric care will be asked to evaluate the completeness of the SAGE. In particular the expertise of the following is required:

- A Geriatrician/expert in rehabilitation

- Assessors of patients unable to answer, nursing home caregivers or family members
- Older adults whom the SAGE was administered.

In respondents and caregivers, Content validity will be assessed through three questions that will be added at the end of the SAGE questionnaire. Questions will ask whether everything expressed in the instrument was clear enough, if some new items needed to be added or the wording needed to be changed.

In case participants, caregivers or experts will individuate sources of confusion or lack of information, the scale will be corrected according to their suggestions. After this process the SAGE will be administered and internal consistency reassessed.

Factor Analysis

Factor analysis is a statistical method used to evaluate the correlation between variables in a new instrument and reduce it to a potentially lower number of non observed variables (or constructs) called factors. This model states that each observed response to an item is partially influenced by underlying common factors. Factors are weighted combinations of all the variables of the scale (intended as items, in our case 15), resulting from the following equations:

$F_1 = w_1X_1 + w_1X_2 + ….. w_1X_{15}$

$F_2 = w_2X_1 + w_2X_2 + ….. w_2X_{15}$

Where X refers to the items and w the weights assigned from the program. The analysis aims to maximize, through a linear correlation approach, the variance existing between variables and condenses the most of it into a number of factors. There are as many factors as items, however weights for the first factor are chosen to account for the most of variance and subsequent factors must account for remaining variance. In the second phase of the factor analysis a rotation of the factor is needed to be able to select factors that are to be retained. Rotation is a mathematical process which permits to discover all sources of variance, that could be hidden by bigger variances [86].

How the number of factors of interest will be determined in SAGE

Usually factor analysis is used to determine how many factors of a scale will be retained. As reported by Hayton and al[87], general methods will be used to decide which factors of SAGE will be retained and which will be discarded are:

- Factor loading matrix

  The factor loading matrix is a table of values, which describes the amount of variance with each item that contributes to explaining the factor. Indeed in the factor loading matrix are included a series of parameters called Eigenvalues, which measures of variance. The number of Eigenvalues higher than zero resulting from the analysis, will indicate the number of factor needed to explain all the variance in the correlation matrix (called the K1 criterion). According to the criterion, firstly introduced by Kaiser and

later confirmed by Guttman[82], only factors with Eigenvalue greater than one will be retained.

- Cattell's Scree test

  The Scree Test is an examination of a plot of eigenvalues to determine discontinuities. The rationale behind the test is that a few major factors usually account for most of variance. The plot will result in a skew as these factors are identified and an "elbow" where the minor factors start to account for the rest of variance. The breakpoint (the elbow) is the sign to look for when trying to identify the number of factors to retain.

Factor analysis will be useful to understand in particular if SAGE is, as supposed, able to measure disability as a unique construct. In this Exploratory Factor Analysis it will be observed, through the K1 rule and scree plot, how many underlying constructs are tapped by the items included in the tool.

CONSTRUCT VALIDITY

Construct validity, as already outlined, describes the extent to which the new scale is really measuring the trait it is supposed to measure. The approach that will be used to assess the relationship between disability and the observable variables related, are box plots and ANOVA test will be used to ensure the significance of the mean differences observed between groups.

Box Plots

In order to assess the trend of SAGE scores on the basis of the different degrees of disability in the three groups of population selected, we will perform an exploratory series of vertical box plots, stratifying the scores by site of recruitment. Box plots are very useful diagrams indicating the median value, the value corresponding to the lowest quartile and the highest quartile. Box plots give also the idea of how many scores fall outside the higher or the lower quartile. This information will be important in the preliminary phase of analysis because outliers are a dispersion index and it is worthwhile to further investigate the reasons why those scores are so different from the rest of the population[82].

Since the population has different degrees of disability it is expected that SAGE would reflect this difference and that the scores would fall accordingly in the range of 0-15 for the community dwelling patients, in the range 16-30 for the ICU patients and in the range 31-45 participants living in a nursing home.

Mean values and standard deviations will be reported by centre and mean differences between the three groups of respondents will be tested for statistical significance with the ANOVA test.

CONCURRENT CRITERION VALIDITY

Concurrent criterion validity of SAGE will be determined by examining the existing correlations between scores of SAGE and the scores of MoCA, FAI, m-Rankin and CSHA Clinical Frailty Scale. We will assume that between the scores obtained with scales measuring different aspects of disability and SAGE (which

comprises all this aspects in one instrument) there will be a linear correlation. In other words we are confident we will be able to prove that as SAGE score decrease, the FAI score and Clinical Frailty scale classification will also decrease, while MoCA and m-Rankin scores will increase in parallel. Pearson's coefficient of correlation is the preferred statistical procedure to assess linear correlation therefore Pearsons' r will be reported overall and by centre.

Pearson's coefficient of correlation

The Pearsons' product moment correlation (also called coefficient r) is the measure of the correlation existing between two continuous variables. The sign and size of the coefficient r give the direction and the magnitude of the relationship between two variables. Advantages of r are that it is relatively useful to partition the variance of the two measures in components that can be meaningful it can also be used to predict one variable from a set of other variables because it is a measure of the linear relationship of the variables of interest.

Pearson's r is defined as the covariance of two variables divided by the product of their standard deviation. The covariance of two variables X and Y, is the average of products of the deviations of single observations X and Y from their respective means.

$$\text{Covariance}_{xy} = \frac{\sum(X-\mu x)\,(Y-\mu y)}{N}$$

$$r = \frac{Cov\,(X,Y)}{\sqrt{Var(X)x\ var\ (Y)}}$$

Interpretation of the r coefficient

The coefficient of correlation sign is the slope of the line expressing the linear correlation and the coefficient itself the degree of linear association existing between the two variables. The Pearsons' correlation coefficient can have values ranging from -1 and + 1. However, caution is needed when considering whether Pearson's correlation coefficients are appropriate or not, because the interpretation depends on the context and the purposes of the study. Usually a correlation coefficient of 0.7 is considered acceptable. However, in context in which some standardized measures have been used, it is desirable to have a higher value considered for acceptance. In table 6 the methods used to assess validity of the SAGE scale is reported.

Table 6. Methods used to assess validity of SAGE scale

Type of validity	Method used
Content Validity	- Expert opinion - Factor Analysis
Construct Validity	- Box Plots - ANOVA test
Concurrent Criterion validity	- Pearson's correlation between: SAGE-MOCA SAGE-FAI SAGE-mRS

	SAGE- CSHA Clinical Frailty Scale

6.7 Hierarchical sequence of the items

SAGE has been developed with the aim to follow the real hierarchical sequence of the loss of independence in the elderly. Therefore we will test the hierarchical properties of the construct by using the Mokken Method. The first author to introduce a theory on the basis of which test hierarchical sequence of items was Guttman, theorizing that a subject giving a positive response to a more difficult item would also give a positive response to a less difficult one[88]. However the Guttman method, which is referred to as the Classical test Theory (CTT), has some disadvantages and some limitations. First of all, in CTT the analysis performed on items and scales are only valid in the group of people who took the test, secondly it is assumed that each item contributes equally to the final score, and that the error of measurement is equal everywhere in the scale (homoscedasticity). In the late 60's, to overcome the CTT's shortcomings, a new theory was developed called Item Response Theory (IRT). IRT is based on two assumptions: items tap only one trait (unidimensionality) and responses to the same item will be different also in participants with the same trait (local independence).

In SAGE the sequence of the items goes from the tasks that require attention, participation and independence to those typical of completely dependent

subjects. We will assess the hierarchical sequence of the items with the Mokken Method described below.

The Mokken method

The Mokken method through a probabilistic approach, measures the probability of answering items based on the level of the trait for each person. The model, as all IRT models, makes three fundamental assumptions:

- Unidimensionality The scale is the reflection of a unidimensional latent trait (in this case the disability). The probability that a subject *a* respond positively to an item (*i*) is indicated by π_{ia}. The probablilty π_{ia} could be quantified by repeated administration of the item *i* to the entire population. Therefore the population could be ordered on the basis of $\pi_{ia..}$ Mokken also defined π_{ia} as a function of the underlying trait (called θ). Therefore π_{ia} and θ are proportionally related $\pi_{ia} = P_i(\theta)$.[89]
- Monotonicity Given that the probability of responding positively to an item increases as the level of θ increases, the probability of responding to a given item should increase monotonically with the higher scores of the scale.
- Local independence Assumes that for any participant with a given value of θ responses will be considered statistically independent.

How the Mokken method measures difficulty

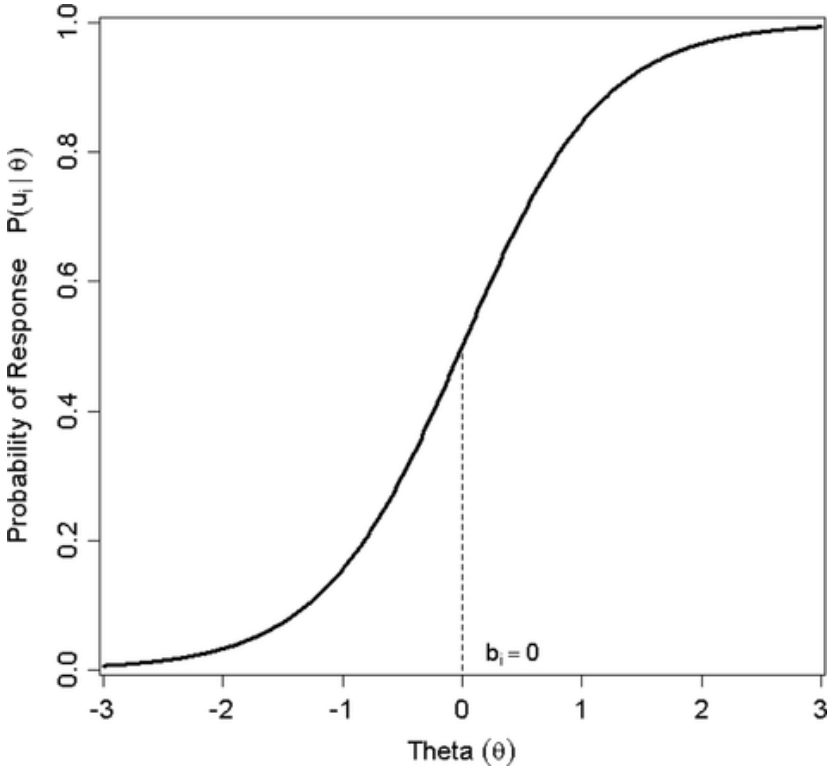
The probability of endorsing one item *i* is related to the difficulty a person with trait θ encounters in answering to it. Mokken defined as b_i the difficulty of an item. To be able to calculate this parameter, he made a new assumption: all participants have to find the same ordered item-related difficulty (Double Monotonicity).

For each item *i* there is a unique value of θ, so that the probability of positively endorsing the item is $P_i(\theta)=0.5$, the difficulty of the item *i* (b_i) is defined as $P_i(b_i)=0.5$ at the same value of θ. Therefore when θ=b, 50 % of the population will endorse the item. While if the $b_i < b_j$ then the probability of endorsing item i will be higher than the probability of endorsing item j.

Testing hierarchical sequence difficulty of the items in SAGE

The probability of endorsing items based on the level of trait (which are participants' characteristics) and on the items difficulty are shown in Item Response Function (IRF) figure 2. Item response functions plot the probability of endorsing an item or a question, over the level of trait for each item or question. The shape of the curve can give an idea of the ability of the question to discriminate different levels of the trait and secondly b will be determined graphically as how much trait is needed to have a probability of positive answers is 50%.

Figure 2 Example of Item Response Function



One assumption of the double monotonicity is that the probability of a positive response will decrease with the item difficulty for any value of θ. In order to assess whether the SAGE scale follows the rules of a Mokken IRT scale, analysis of the monotonicity of the items and the monotonicity in item difficulty will be performed. The coefficient of scalability (H) will also be reported and, as indicated by Mokken, a scalability coefficient greater than 0.5 will be considered proof of hierarchically sequenced items. Previous work already outlined the statistical procedure to follow in order to obtain an analysis of the items difficulty following the Mokken method[90].

Limitation of the Mokken Method

This method is based on dichotomous responses, which have to increase with the theoretical increase of the trait level. Therefore in SAGE we will test as unique trait disability, considering as positive moderate and severe answers to each item comprised in the scale. In this way we will be able to test difficulty for each item.

6.8 Sensitivity and specificity

Sensitivity measures the proportion of positive responses that are correctly identified as positive, while specificity is measured by the proportion of negatives responses that are correctly identified with our new scale.

In our case we would like to prove that SAGE is sensitive in detecting older adults with a disability who are truly disabled, and specific to the issue of disability only. To be able to measure sensitivity and specificity we will need to define numerically disability and non disability. As already reported mRankin and MoCA are very widely used scale to detect the grade of dependence and cognitive ability in the elderly. In previous studies (UK TIA) an mRankin score of 3 was associated with the presence of disability, while MoCA demonstrated to be able to discriminate between subject with no cognitive impairment (score > 26) and those with some grade of cognitive disability (score <26).

To be able to discriminate disabled subjects from those who are not, we will also need a threshold in the SAGE scores. The results of contruct validation will help us decide what is the median score of SAGE that describes a completely

disabled older adults (median values of the subjects recruited in the nursing home facility) or community dwelling older adults (subjects recruited from the HOPE-3 study). On the basis of the SAGE, mRankin and MoCA cut-offs, the cut points for 2 x 2 cross tables identifying true disabled patients scores will be reported as follows:

Determining disabled patients

	Sage score: Disabled	Sage score Able
M Rankin ≥ 3	A= True positive	B=False positive
M Rankin < 3	C=False negative	D=True negative

	Sage score: Disabled	Sage score Able
MoCA ≤ 26	A=True positive	B=False positive
MoCA < 26	C=False negative	D=True negative

In general Sensitivity = $\frac{A}{A+C}$ and

Specificity= $\frac{D}{B+D}$

For each SAGE score will be reported the number of cases correctly identified from the new scale and the number of non cases. In the same table the value of sensitivity and specificity will be enlisted. Methods used to graphically plot the sensitivity over the complement of specificity are the Receiver Operating Characteristic Curves. These curves are useful to understand if the new scale is able to discriminate between people with the disease and people without

Receiver Operating Characteristic Curves (ROC curves)

The Roc Curves are a very useful approach to determine the cut off score able to discriminate between a group of people who are the subjects affected by the disease and the group who are not.

In ROC curves sensitivity is plotted against the complement of specificity (1-Specificity) for each point of score. For example in the SAGE case, we will calculate sensitivity and specificity of the total score ranging from 0 which is a score that could be obtained in a very independent subject to 48, which is the maximum score that can be obtained in a totally disabled and dependent subject. For each point of score on the basis of the true positive and false

positive cases of disability detected by SAGE (obviously compared to the cases and false cases detected by MoCA and mRankin), sensitivity and specificity will be numerically calculated and plotted in a ROC curve. A good discriminator test will have high sensitivity and high specificity and therefore graphically will be described from a curve with an elbow as close to the top left corner as possible.

Limitations of the study

The design of this study has some limitations. The cross-sectional design does not allow any causal-effect conclusion since we will be able to assess participants' ability degree only at one point in time. Moreover it will not be possible to assess inter-rater and test-retest reliability, these will be objectives of a future prospective study on SAGE validity and reliability. In the validation process described in this thesis the use of proxy information will be allowed. Although some studies showed a high agreement between patients-reported and proxy-reported ADL, the use of proxy information is not always reliable. Situations of frailty, cognitive impairment or depression could influence participants' ability or participation in social life and therefore confound their responses.

Conclusions

In this chapter the methodological aspect of the SAGE validation study have been described. The results of this study, will give information on reliability, validity specificity and sensitivity of the new scale. The expected hierarchical sequence of the items in SAGE will also be tested with the Mokken Method. This preliminary phase will be the basis for further studies for example on the predictive validity of SAGE that will be described in more details in the next concluding chapter.

Chapter 7 Conclusions and Future Directions

7.1 Conclusions on the thesis

In the past decades the term “disability” was mostly related to the inability to perform some basic (such as bathing or toileting or preparing meals) and instrumental activities (such as using transportation or managing finances), necessary to live independently. With the recent (2001) recognition of the “burden” of disability as an interplay of different aspects affecting the subject (individual, external or social), a new instrument capable of capturing all these aspects was found to be needed. In the previous chapters, a review of the most widely used instruments for disability assessment validated in the elderly has been reported. Some scales have been developed on the basis of the new definition of disability, but none of them included cognition as part of the measurement (which sometimes could drive the respondent answers). Moreover the characteristics of those constructs, principally length and wording, make them not really applicable in an older adult setting.

The SAGE scale has been developed to focus on the activities that an older adult can complete at a personal, home and at a social level. It consists of 15 items and it’s easily administrable in 10 minutes. However the SAGE scale cannot be administered as it is, since it needs to be validated in a population of older adults. In this thesis, methodological aspects of the validation process of the SAGE scale have been discussed. The results of the validation study will be

useful to determine firstly, whether the use of the SAGE scale and the scores obtained really reflects the independence/dependence status of the subjects assessed. Secondly, validation is also important to assure that the clinician's conclusion based on the scores obtained through SAGE will be correct and reliable.

The SAGE scale could be very helpful, if the study described in this thesis proves its validity, in early detection of disability and loss of independence in older adults. As we know, human life span is increasing and in this context, the SAGE scale could be helpful in discriminating older adults who are still able to live normally in the community from those who need more skilled assistance.

7.2 Future Directions

The results of the SAGE's validation study will be useful to determine whether further studies will be performed. If the SAGE instrument is proved to be a valid tool in detecting disability in older adults, a prospective validity study should be designed in a much larger, multi ethnic population. The aims of this study are firstly, to assess test retest reliability by administering the scale in two different points in time. Secondly, it would essential to test the predictive validity of the tool. Predictive validity is the extent to which a score or a scale can predict some events or scores of a criterion measure in the future. For this reason the SAGE scale could be administered at one point in time and the events recorded or the criterion measure administered in a second point in time.

The opportunity will be given by the APOLLO study: an international randomized, double blind, factorial controlled trial on the efficacy of Aliskiren (a new antihypertensive drug) in elderly subjects. The APOLLO study recruitment started in February 2011 and is going to include 11, 000 subjects. During 5 years of follow up the efficacy of Aliskiren, in combination with other antihypertensive drugs (such as diuretics or calcium channel blockers), will be tested on major cardiovascular events, but also on the participants' ability to carry on the activities associated with independent living. In this context SAGE could be administered to a subsample of the study population to evaluate predictive validity. The APOLLO study will be also the perfect opportunity to test the SAGE scale in elderly population from different countries.

7.3 Conclusions

The thesis described the methodology and the theory of the validation of a new tool for the assessment of disability in the older adults: the SAGE scale. The development of this tool was motivated from the recent need of an instrument able to capture all the activities that are important for the elderly to be able to age with dignity and independence. The results of this study, if positive, will be useful for further investigation of the SAGE and its implementation as a screening tool to recognize and detect early loss of independence in this group of individuals.

Bibliography

1 Verbrugge LM, Jette AM. The disablement process. Soc Sci Med, 1994; 38: 1-14.

2 Gill TM, Robison JT, Tinetti ME. Difficulty and dependence: two components of the disability continuum among community-living older persons. Ann Intern Med 1998; 128: 96-101.

3 World Health Organization. International Classification of Functioning Disability and Health, Geneva: WHO; 2001

4 Whiteneck G, Dijkers MP. Difficult to measure constructs: Conceptual and Methodological issues concerning participation and eviromental factors. Arch Phys Med Rehabil 2009; 90 (11 Suppl 1): S22-35.

5 Desrosiers J, Robichaud L, Demers L, Gèlinas I, Noreau L, Durand D. Comparison and correlates of participation in older adults without disabilities. Archives of Gerontology and Geriatrics 2009; 48: 397-403.

6 Wilkie R, Peat G, Thomas E, Croft P. Factors associated with participation restriction in community dwelling adults 50 years and over. Qual life Res, 2007; 16: 1147-1156.

7 Millàn-Callienti JC, Tubio J, Pita-Fernandez S, Rochette S, Trinidad L, Ana M. Cognitive impairment as predictor of functional dependence in an elderly sample. Archives of gerontology and geriatrics, 2011; http://dx.doi.org/10.1016/j.archger.2011.02.010

8 Von Strauss E, Viitanen M, De Ronchi D et al. Aging and the occurrence of dementia: findings from a population-based cohort with a large sample of nonagenarians. Arch Neurol 1999; 56: 587-592.

9 Romàn G, Erkinjuntti T, Wallin A, Pantoni A, Chui CH. Subcortical ischaemic vascular dementia. Lancet Neurology 2002; 1:426-36.

10 Forette F, Seux ML, Staessen JA, Ltgarde T et al. Prevention of dementia in randomized double blind placebo-contolled Systolic Hypertension in Europe (Sys-Europe) trial. Lancet 1998; 353 (9137): 1347-51.

11 Van Dijk E, Prins ND, Vrooman HA, Hofman A, Peter J et al. Progression of cerebral small vessel disease in relation to risk factors and cognitive consequences: e Rotterdam Scan study. Stroke, 2008; 39: 2712-2719.

12 Pedelty L, Gorelik P. Management of hypertension and cerebrovascular disease in the elderly. Am J Med 2008; 121 (8) suppl 1: 23-31.

13 Qiu C, Winblad B, Fratiglioni L. The age-dependent relation of blood pressure to cognitive function and dementia. Lancet Neurology 2005; 4:487-99.

14 Ikram MA, van Oijen M, de Jong J, Kors JA, et al. Unrecognized Myocardial Infarction in relation to risk of dementia and cerebral small vessel disease. Stroke 2008; 39: 1421-1426.

15 Ott A, Stolk RP, van Harskamp F, Pols H.A.P et al. Diabetes Mellitus and risk of dementia, the Rotterdam Study. Neurology, 1999; 53: 1937-1942.

16 Roberts RO, Geda YE, Knopman DS, Christianson TJH et al. association of duration and severity of Diabetes Mellitus with Mild cognitive Impairment. Archives of Neurology 2008; 65(8): 1066-1073.

17 Bergman H, Ferrucci L, Guralnik J, Hogan DB, Hummel S, Karuanananthan S, Wolfson C. Frailty: An Emerging and Clinical Paradigm-Issues and Controversies. J Gerontol A Biol Sci Med Sci, 2007; 62(7): 731-737.

18 Abellan Van Kan G, Rolland Y, Houles M, Gillette-Guyonnet S, Soto M, Vellas B. The assessment of frailty in older adults. Clin Ger Med, 2010; 26: 275-286.

19 Rowe J, Kahn R. Human Ageing: Usual and Successful. Science 1987(7); 143-150

20 Depp CA, Jeste VD. Definitions and predictors of Successful Aging; a comprehensive review of larger quantitative studies. Am J Geriatr Phychiatry 2006; 14(1): 6-20.

21 Reichstadt J, Depp CA, Palinkas LA, Folsom DP, Jeste DV. Building blocks of successful aging: a focus group study of older adults' perceived contributors to successful aging. Am J Geriatr Psychiatry, 2007; 15(3): 194-201.

22 Phelan EA, Anderson LA, LaCroix AZ, Larson BE. Older Adults' Views of ''Successful Aging'' How Do They Compare with Researchers' Definitions? J Am Geriatr Soc 2004; 52:211–216.

23 Horner B, Boldy DP, The benefit and burden of “ageing in place” in an aged care community. Australian Health Review; 2008 32(2): 356-365.

24 Alley D, Liebig P, Pynoos J, Banerjee T, Choi H. Creating Elder-Friedly Communities: preparations for an Ageing Society. Journal of Gerontological social Work, 2011; 49:1-2, 1-18.

25 McDonough K, Davitt KJ. It Takes a Village: Community Practice, Social Work and Ageing in Place. Journal of Gerontological social Work, 2011; 54:5 528-541.

26 Griffith L, Raina P, Wu H, Zhu B, Stathokostas L. Population attributable risk for functional disability associated with chronic conditions in Canadian Older adults. Age Ageing, 2010;39(6):738-45

27 Sousa RM, Ferri CP, Acosta D, Albanese E, Guerra M, Huang Y et al. Contribution of Chronic diseases to disability in countries with low and middle incomes: a 10/66 Dementia Research Group population based survey. Lancet, 2009; 374 (9074); 1821-1830

28 Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020. Lancet, 1997; 349 (9064): 1498-504

29 Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW, Cleveland MA. Studies of illness in the aged. The index of ADL: a standardized measure of Biological and Psychosocial Function. JAMA, 1963; 185: 914-919.

30 Wiener JM, Hanley RJ, Clark R, Van Nostrand JF. Measuring activities of daily living: Comparisons Across National Surveys. Journal of Gerontology, 1990; 45(9): S229-237.

31 Jagger C, Arthur AJ, Spiers NA, Clarke M. Patterns of Onset of Disability in Activities of Daily living with Age. J Am Geriatr Soc, 2001; 49: 404-409.

32 Brorsson B, Asberg KH. Kats index of independence in ADL. Reliability and validity in short-term care. Scand J Rehabil Med, 1984; 16 (3): 125-32.

33 Wallace M. Katz Index of Independence in Activity of Daily Living (ADL). Annals of long-term Care, 2006, Issue 11. Online version

34 Reijneveld SA, Spijker J, Dijkshoorn H. Katz' ADL indec assessed functional performance in Turkish, Moroccan and Ditch elderly. Journal of clinical Epidemiology, 2007; 60 (4): 382-388

35 Mahoney FI, Barthel D. Functional evaluation: The Barthel Index. Md State Med J, 1965; 14: 56-61.

36 Wylie CM. Measuring end results of rehabilitation of patients with Stroke. Pub Health Rep, 1967; 82: 893-898.

37 Shah S, Venclay F, Cooper B. Predicting discharge status at commencement of stroke rehabilitation. Stroke, 1989; 20: 766-769

38 De Morton N, Keating JL, Davidson M. Rasch Analysis of the Bartel Index in the Assessment of Hospitalized Older Patients after admission for an Acute Medical Condition. Arch Phys Med Rehabil, 2008; 89: 641-7.

39 Lawton M, Brody ME. Assessment of older people: Self Maintaining and Instrumental Activities of Daily Living. Gerontologist, 1969; 9: 179- 186.

40 Van Der Putten JJMF, Hobart JC, Freeman JA, Thompson AJ. Measuring change in disability after inpatients rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. J Neurol Neurosurg Psychiatry, 1999; 66: 480-484.

41 Hsueh IP, Lin JH, Jeng JS, Hsieh CL. Comparison of the psychometric characteristics of the functiona independence measure, 5 item Barthel index, and 10 item Barthel index in patients with stroke. J Neurol Neurosurg Psychiatry, 2002; 73: 188-190

42 Pollak N, Rheault W, Stoecker JL. Reliability and Validity of the FIM for persons aged 80 years and above from a multilevel continuing care retirement community. Arch Phys Rehabil, 1996; 77: 1056- 61.

43 Spector W, Katz S, Murphy J, Fulton JP. The hierarchical relationship between activities of daily living and instrumental activities of daily living. J Chron Dis, 1987; 40(6): 481-89.

44 LaPlante MP. The classic measure of disability in Activities of Daily living is biased by age but an expanded IADL/ADL measure is not. Journal of Gerontology, 2010; 65(6): 720-732.

45 Holbrook M, Skilbeck CE. An activity index for use with stroke patients. Age and Ageing, 1983; 12(2): 166-170.

46 Wade DT, Legh-Smith J, Langton Hewer R. Social activities after stroke: measurement and natural history using the Frenchay Activity Index. Int. Rehabil. Med, 1985; 7(4): 176-181.

47 Wan Han C, Joo Lee E, Kohzuki M. Validity and reliability of the Frenchay Activity Index in community dwelling elderly in South Korea. Tohoku J Exp Med, 2009; 217 (3) 163-68.

48 Clemson L, Bundy A, Unsworth C, Singh M. Validation of the modified assessment of living skills and resources, a IADL measure for older people. Disability and Rehabilitation, 2009; 31(5): 359-369.

49 Zimnavoda T, Weinblatt N, Katz N. Validity of the Kholman evaluation of Living Skills with Israeli elderly individuals living in the community. Occupational Therapy International, 2002; 9(4): 315-325.

50 Brunett J, Dyer CB, Naik AD. Convergent validation of the Kholman Evaluation of living skills as a screening tool of older adults' ability ti live safely ad independently in the community. Arch Phys Med Rehabil, 2009; 90; 1948-52.

51 Haley SM, Jette AM, Costner WJ, Kooyoomjian JT, Levenson S, Heeren T, Ashba J. Late Life Function and disability Instrument; Development and evaluation of the function component. J Gerontol A Biol Sci Med Sci, 2002; 57(4): M 217-22.

52 Jette AM, Haley SM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, Ashba J. Life Function and disability Instrument; Development and

evaluation of the disability component. . J Gerontol A Biol Sci Med Sci, 2002; 57(4): M 209-16.

53 Sayers SP, Jette AM, Haley SM, Hereen TC, Guralnik JM, Fielding RA. Validation of the Late Life Function and disability Instrument. Journal of the American Geriatric Society, 2004; 52: 1554-1559.

54 Jette AM. Towards a common language of disablement. J Gerontol A Biol Sci Med Sci, 2009: 64A(11) 1165-68.

55 Freedman VA. Adopting the ICF language for studying Late-life Disability: a field of dreams? J Gerontol A Biol Sci Med Sci, 2009: 64A(11) 1172-74.

56 Grill E, Stucki G, Boldt C, Joisten S, Swoboda W. identification of relevant ICF categories in geriatric patients in an early post-acute rehabilitation facility. Disability and rehabilitation, 2005; 27(7,8): 467-73

57 Grill E, Joisten S, Swoboda W, Stuki G. Early stage impairmaints and limitation of functioning from the geriatric IGF corset As determinants of independent living in older patients after discharge. J Rehabil Med, 2007; 39: 591-597.

58 Üstün BT, Chatterij S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J. developing the World Health Organization Disability Assessment Schedule 2.0. Bull World Health Organ, 2010; 88: 815-823.

59 Rehm J, Üstün BT, Shekar S, Nelson C, Somnath C. On the development and psychometric testing of the WHO screening instrument to assess

disablement in the general population. International Journal of Methods in Psychiatric research, 1999; 8: 110-122.

60 http://www.who.int/classifications/icf/WHODAS2.0_12itemsSELF.pdf

61 Sousa RM, Dewey ME, Acosta D, Jotheeswaran AT, Castro-Costa E, Ferri PC. Measuring disability across cultures-the psychometric properties of the WHODAS II in older people from low- and middle-income countries. The 10/66 Dementia Research Group population-based survey. International Journal of Methods in Psychiatric research, 2010; 19(1): 1-17.

62 Hobart, J.C. & Thompson, A. J."The five item Barthel index", J.Neurol. Neurosurg. Psychiatry, 2001; vol. 71, no. 2, pp. 225-230.

63 Graf, C. "The Lawton instrumental activities of daily living scale", Am.J.Nurs., 2008; vol. 108, no. 4, pp. 52-62

64 Folstein MF, Folstein S, McHugh P, “Mini Mental State”: A Practical Method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 1975; 12 (3): 189-198.

65 Normann G, Strainer D. Health Measurements Scales: A Practical guide to their development and Use. Oxford Medical Publications, 3rd edition. Pp 106.

66 http://www.who.int/topics/disabilities/en/

67 Norman G, Strainer D, Health Measurements Scales: A Practical guide to their development and Use. Oxford Medical Publications, 3rd edition. Chapter 10

68 Landy F.J. Validation as Hypothesis Testing. American psychologist, 1986; 41: 1183-92

69 Chronbach L.J, Meehl P.E. Construct validity in psychological tests. Psychological Bulletin, 1955; 52: 281-302.

70 Kane M.T, Current concerns in validity Theory. Journal of Educational Measurement, 2001; 38: 319-342

71 Messick S. Validity of Psychological Assessment. American psychologist, 1995; 50: 741-749.

72 Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical research, 3rd edition, 2007, Lippinkott Williams and Wilkins, Chapter 8.

73 Nasreddine Z, Philipps N, Bèdirian V, Charbonneau S, Whitehead V, Collin I. The Montreal Cognitive Assessment MoCA: A brief screening tool for Mild cognitive impairment. Journal of the American Geriatric Society, 2005; 53: 695-699

74 Koski L, Xie H, Finch L. Measuring cognition in a geriatric outpatient clinic: Rasch analysis of the Montreal Cognitive Assessment. Journal of geriatric psychiatry and neurology, 2009; 22(3): 151-160.

75 Quinn TJ, Dawson J, Walters M. Dr John Rankin; His lihe, legacy and the 50th Anniversary of the Rankin Stroke scale. Scott Med J, 2008; 53(1): 44-47.

76 UK-TIA Study group. The United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: final results. Journal of Neurology, Neurosurgery and Psychiatry, 1991; 54: 1044-54.

77 Banks JL, Marotta CA. Outcomes Validity and reliability of the Modified Rankin Scale: Implications for stroke clinical trials: a literature review and synthesis. Stroke, 2007; 38: 1091-96.

78 Rockwood K, Xiaowei S, MacKnight C, Bergman H, Hogan DB, McDowell I, Mitnitski A. A global clinical measure of fitness and frailty. Canadian Medical Association journal, 2005; 173(5): 489-495.

79 Oczkowsky C, O'Donnell M. Reliability of proxy respondents for patients with stroke: A systematic review. Journal of Stroke and Cerebrovascular Diseases, 2010. 19(5) 410-16

80 http://www.mocatest.org/pdf_files/MoCA-Instructions-English_2010.pdf

81 http://www.mocatest.org/validation_study.asp

82 Strainer D, Norman G. Biostatistics, Bare essentials . BC Decker, 3rd edition. 2008

83 Strainer D, Norman G. Health Measurement Scales. Oxford Unversity Press, 3rd edition 2008, chapter 4.

84 Kane MT Interpreting variance components as Evidence for Reliability and Validity, Paper presented at the annual Meeting of the American Educational Research Association April, 1980 pg 1-61

85 Ponterotto J, Ruckdeschel DE, An overview on coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. Perceptual and Motor Skills, 2007; 105:997-1014.

86 Nunally, Bernstein. Psychometric Theory. 1994, 3rd edition McGraw-Hill, p 470.

87 Hayton JC, Allen DG, Scarpello V. Factor retention decisions in Exploratory Factor Analysis: A tutorial on parallel analysis. Organizational research methods 2004; 7: 191-205.

88 Guttman L. Some necessary conditions for common factor analysis, psychometrika 1954; 19: 149-162

89 Mokken R, Lewis J. A non parametric approach to the analysis of dichotomous item responses, Applied Psychological measurements, 1982; 6 (4)417-430.

90 Kingma J, Taerum T. SPSS-X procedure and stand alone programs for the Mokken Scale Analysis: a non parametric item response theory model. Educational and psychological Measurement, 1989; 49: 101-136.