Identifying longitudinal sustainable hierarchies in activities of daily living

Running title: ADL hierarchies

Roger Watson*
Faculty of Health & Social Care
University of Hull, Hull HU6 7RX, UK
e-mail: r.watson@hull.ac.uk

Alvisa Palese
Department of Medical and Biological Sciences
Udine University, Italy
e-mail: alvisa.palese@uniud.it

Ranieri Zuttion
Welfare Unit
Friuli Venezia Region, Italy
e-mail: ranieri.zuttion@welfare.fvg.it

Barbara Ferrario
Welfare Unit
Friuli Venezia Region, Italy
e-mail: barbara.ferrario@welfare.fvg.it

Sandra Ponta
Friuli Venezia Region, Italy
e-mail: sandra.ponta@welfare.fvg.it

Mark Hayter
Faculty of Health & Social Care
University of Hull, UK
e-mail: m.hayter@hull.ac.uk

* corresponding author

Key words: Mokken scaling, item response theory, disability, older people, nursing homes, Italy, occupational therapy, nursing, Val.Graf instrument

ABSTRACT

Activities of daily living serve as an indicator of progression in disability and rehabilitation. It is know that some of the measurement scales used show hierarchical properties indicating that activities of daily living are lost and gained in a consistent pattern. Few studies have investigated the extent to which these patterns are sustained across time and across a range of disability. The study aimed to investigate the hierarchical properties of the activity of daily living items in the ValGraf functional ability scale, to establish if there is a hierarchy of items in the scale and to study the sustainability of the hierarchy over time. Secondary analysis of a retrospective database from 13,113 people over 65 years in 105 nursing homes in northern Italy, between 2008 to 2013 was conducted. Data were gathered 6-monthly and analysed using Mokken scaling to identify a hierarchy of items in the scale and if this was sustainable over time. A sustainable hierarchy of items was observed running in difficulty from urinary incontinence to feeding. The hierarchical structure of the activities of daily living observed in the present study is stable over time meaning that changes in total score for these items can be compared meaningfully across time.

Key words:

Mokken scaling

Activities of daily living

Nursing homes

ValGraf instrument

1. INTRODUCTION

Activities of daily living (ADLs) are those activities that we take for granted in health such as walking, climbing stairs and eating. However, in ill health and disability, activities of daily living can become impaired and this often indicates the need for assistance, sometimes in the form of rehabilitation or long-term care. Therefore, ADLs are assessed in healthcare by nurses, physicians and occupational therapists as proxy performance measures of disability.

Several measures of ADL exist and prime examples are the Katz Index of Independence in Activities of Daily Living and the Barthel Index (Katz, 2003). There is considerable similarity between these scales and, therefore, considerable overlap between the items they contain. ADLs are distinguished from what are known as instrumental activities of daily living (IADLs) which are more neurologically complex activities we can carry out provided ADLs are intact such as shopping and using the telephone. Intact IADLS allow us to live independently. Some scales focus exclusively on ADLs and some, such as The Lawton Instrumental Activities of Daily Living Scale (Lawton and Brody, 1969), focus exclusively on IADLs and some, such as the Townsend Activities of Daily Living Scale (MRCCFAS, 1998), cover both domains. Loss of IADL ability may be predictive of disability and followed by loss of ADL activity (Feio et al., 2009) but the distinction between the two is not exclusive. Loss of IADLs may be compensated for by some support in the community. However, loss of ADL function usually necessitates some kind of institutional care which may be focused on rehabilitation, aimed at improving ADL function, or on continuing care as ADL function declines. The present study is focused exclusively on ADL decline in long-term care of older people.

Functional ability as expressed by ADLs is known to decline over time in older nursing home residents (Palese et al., 2016) and it has long been known that ADLs decline in an hierarchical pattern (Gerrard, 2013). In other words, there is a pattern to the decline and retention of ADLs meaning that when ADLs decline some ADLs are consistently lost before others. For example, fine motor movements such as those required for dressing are lost later than those requiring more gross motor movement such as walking (Daltroy et al., 1992; Dunlop et al., 1997). These patterns can be deduced from observing individuals and groups. At the group level these patterns are expressed in the mean values for individual ADLs. Assuming that higher scores mean greater disability, items that are 'difficult' to accomplish have high mean values and those that are less easily accomplished—the 'easiest' items—have low mean values. The extent to which these hierarchies have been formally established varies and the population in which such hierarchies are found is older adults, the majority of whom are from community or acute care settings (Gerrard, 2013). Nevertheless, due to the utility of establishing such hierarchies, and the persistent utility of measuring ADLs as well as actual performance (Kruse et al., 2013; Hjaltadóttir et al., 2012), the issue remains worthy of study. In nursing home facilities it can also inform policy decision-makers about re-designing care priorities and services (Kruse et al., 2013; Hjaltadóttir et al., 2012).

Various methods are available to study the hierarchical properties of ADL scales and these fall within a broad group of methods referred to as item response theory (IRT). Whilst the label did not apply at the time, these methods are derived from Guttman scaling (Stouffer et al., 1950). Guttman scaling was originally used to study the performance of individual items in scales as opposed to methods derived from classical test theory (CTT) (van Schuur, 2003). These methods include Cronbach's alpha and factor analysis, which are mainly—although not exclusively—concerned with test level information. Guttman scaling was used to study whether dichotomous items were ordered in scales and has been applied to activities of daily living scales (Daltroy et al., 1992; Lazardis et al., 1994). A scale is ordered—i.e. it has a hierarchy of items—if items are scored consistently relative to one another. For example, if an item ‘able to walk’ was more readily endorsed than ‘can climb stairs’ and this was consistent across a group of respondents then those items would be ordered in terms of difficulty with ‘can climb stairs’ being more difficult than ‘able to walk’.

Guttman scaling is limited in two ways: first, it is only capable of analysing dichotomous items. For polytomous items—e.g. as applied to the early studies of the Edinburgh Feeding Evaluation in Dementia (EdFED) scale (Watson, 1996)—trichotomous variables had to be dichotomised for analysis with some loss of information about the items in terms of range and responsiveness. Second, the method is deterministic in that it assumes a strict hierarchy of responses (i.e. someone endorsing ‘can climb stairs’ will always endorse ‘able to walk’) and assesses scales merely on the basis of percentage conformity with the proposed hierarchy. More sophisticated IRT methods have been developed which are stochastic in nature and allow for a probabilistic relationship between response patterns (i.e. a person endorsing ‘can climb stairs’ is more likely to endorse ‘able to walk’) and putative hierarchies. The most common IRT methods are based on Rasch scaling (Meijer et al., 1990), a parametric method which, in common with other parametric methods, is capable of creating interval level measurements. These methods are very demanding of data and also often unnecessary in the study of hierarchies where the mere ordering of items provides useful and sufficient information. This was the basis of an alternative, and increasingly popular non-parametric method of IRT, Mokken scaling.

Few studies of ADL have investigated sustainability of item hierarchies over time. In fact, a search of the highly inclusive Google Scholar yielded only the study from the Aging in Manitoba Longitudinal Study which, using Rasch analysis (Finlayson et al., 2005), demonstrated that the ordering of items in the instrument used in this study was stable over three time periods (1983, 1990, 1996) and two different settings (home, nursing home) in a sample of 607 older people. More recently, a hierarchy in the Katz ADLs index was established using Rash analysis, using a cross-sectional data set of 13,113 participants living in skilled nursing facilities (Gerrard, 2013). Rasch analysis and other IRT methods were used to demonstrate a hierarchy of items in the Barthel Index of ADL in 788 long-term care residents (Liu et al., 2015). Clearly, if an ordering to the endorsement of items in an ADL scale is assumed based on cross-sectional data, and this is assumed to indicate the order in which people become disabled, then the extent to which the same ordering of items is apparent over time is an important phenomenon. Otherwise, the predictive value of an ADL scale is compromised; without a sustainable hierarchy of items an ADL scale is merely a tool which is used repeatedly to provide a total score which indicates some overall decline in ability. However, the information ascribed to the scores over time is less meaningful. If ADL scales have predictive ability then they indicate more precisely which losses are contributing to loss of function and they help caregivers to predict need for care and which aspects of self-care deficits to target.

1.1 The present study

A unique opportunity arose to study ADLs data from a large longitudinal sample (N > 10,000) of older people admitted to nursing homes in Italy. Two main research questions guided the study:

1. Is there a Mokken scale in these ADL data?
2. Is the ordering if items in the ADL data stable over time?

In addition, if there is a Mokken scale in the data, the following questions is asked:

3. How stable are Mokken scaling parameters over time?

2. MATERIALS AND METHODS

This is a longitudinal retrospective study involving secondary data analysis. Data for this study were obtained from people over 65 years old in 105 nursing homes in northern Italy, between 2008 - 2013. Data were collected by nurses trained in the use of the multidimensional geriatric assessment ValGraf instrument (Pascazio et al., 2009) which includes items from the Functional Autonomy Measurement System (Hebert and Bilodeau, 1988) and the Barthel Index (Mahoney and Barthel, 1965). The utility and factorial and concurrent validity of the ValGraf instrument are established (Pascazio et al., 2009).

The present study focuses on 13 ADLs shown in Table 1. Data were collected from participants every 6 months and participants were included if they had at least two assessments at admission and 6 months after admission to their nursing home; they were discharged from the nursing home at death or transfer in to another nursing home. Ethical approval was obtained from the Ethic Committee of Azienda Ospedaliero-Universitaria (Udine, Italy) and data were made anonymous prior to analysis.

Data were transferred for cleaning and descriptive and inferential analysis (t-test; analysis of covariance (ANCOVA); repeated measures ANOVA) into an SPSS version 22.0 database

(Supplementary file 1) and then converted using the package 'foreign' in *R* (Van der Ark, 2014) into a format suitable for analysis using the package 'mokken' in *R* (Van der Ark, 2014) (Supplementary file 2).

2.1 Mokken scaling

Mokken scaling analysis (MSA) is a method within the field of item response theory (IRT) which is non-parametric (Mokken and Lewis, 1982). Unlike parametric IRT methods, such as Rasch scaling, Mokken scaling makes fewer assumptions about the shape of the item characteristic curve (ICC). The ICC describes the relationship between a person's position on a latent trait and the person's score on an item measuring that latent trait. The advantage of a non-parametric model such as MSA is that it tends to be more inclusive of items. All that is required is that ICCs continuously increase over the latent trait being measured—are monotonous— and that ICCs do not intersect (Roskam et al., 1986). In the case of items with polytomous response categories. This is called invariant item ordering (IIO) (Sijtsma et al., 2011). The value of IRT methods over methods such as factor analysis is that IRT methods more accurately relate the total score on a trait to the particular items that have been scored as an ordering in the way respondents score items is assumed (Watson et al., 2012). Mokken scales are assumed to be unidimensional and items are assumed to be locally stochastically independent (Watson et al., 2014).

A range of parameters is used to judge whether a set of items forms a Mokken scale. Foremost amongst these is Loevinger's coefficient (*H*) which is a measure of the scalability of a set of items (Mokken and Lewis, 1982). Mokken scaling is a stochastic development of Guttman scaling and Guttman errors—where a pair of items is not scored in the expected direction—are used to calculate Loevinger's coefficient and, thereby, to judge the strength of a Mokken scale. Higher values of *H* represent stronger scales (>0.3 is a weak scale; >0.4 is a moderate scale; >0.5 is a strong scale) and Loevinger's coefficient can also be calculated for items (*Hi*) and pairs of items (*Hij*). By calculating the standard errors of these coefficients, confidence intervals can be calculated (Kuijpers et al., 2013) to examine whether items or pairs of items should be retained in a set of items to form a Mokken scale. For *Hi* the 95% confidence intervals should not include 0.3 and for *Hij* the 95% confidence intervals should not include 0. Monotone homogeneity can be investigated using a value called *Crit* (van Schuur, 2003) which is generated in MSA software by taking into account the frequency and seriousness of violations of monotone homogeneity and values of *Crit* below 40 are considered to be acceptable for items in a Mokken scale. IIO is judged by the distance between items along the latent trait and this is judged by calculating *H*trans (denoted H^T) (Sijtsma et al., 2011) which is analogous to *H* and the values at which weak, moderate and strong IIO is judged are the same as for *H*. However, values of H^T alone should not be used to judge IIO. It has been shown that individual ICCs positioned far along the latent trait from the remainder can lead to misleadingly high values of H^T and exaggerate IIO (Meijer and Egberink, 2012). Therefore, it is advisable to plot ICC pairs and to examine them for 'outliers' from the main cluster of ICCs and to recalculate H^T in the absence of these items. Scale reliability (*Rho*) can be calculated and the probability of obtaining a Mokken scale can be estimated using a Bonferroni method which corrects for multiple iterations.

The number of putative Mokken scales at each data collection point was determined using the automated item selection procedure. The following parameters were also determined using the default settings in *R* package 'mokken': *H*, monotonicity; H^T, IIO and significant violations of IIO (syntax is given in Supplementary file 3). Standard errors of *Hij* were only studied at baseline (a previous unpublished study—available on request—shows that this is relatively insensitive to sample size, therefore this was not followed up over the remainder of the study). The activities of daily living entered into the MSA are shown in Table 1. The item scores ranged from 1-5 (low to high dependency) for all items except two related to bladder and bowel continence which ranged from 1-6. The number of people scoring 6 on these items was in single figures, therefore they were removed from the database thereby meaning that the score range for all items analysed by Mokken scaling was the same. A sensitivity analysis comparing males and female scores was also conducted.

3. RESULTS

The demographics of the sample were as follows: males (n=2939; mean age 78.28 SD 10.86 years); females (n=8803; mean age 85.20 SD 8.47 years). At all 11 time points in the study only a single Mokken scale was found in the data with no violations of monotonicity. At baseline none of the 95% confidence intervals for *Hij* included 0. Across the whole study the ordering of the items by mean score was maintained (Appendix 1). Figures 1 – 5 show several ICC pair plots. Figures 1 and 2 exemplify that the ICC of one of the items entered into the initial analysis—appetite disturbance/nutrition—was positioned far away from the remaining cluster of items and showed no discrimination between individuals due to a floor effect. Therefore, the item 'nutrition' was omitted from the remainder of the analysis. Figure 3 shows two items that have intersecting ICCs and which, therefore, do not show IIO and Figures 4 and 5 show two pairs of items with non-intersecting ICCs. Based on a visual inspection of all the pairs of ICCs (Supplementary file 4), with the exclusion of the item 'appetite disturbance/nutrition', the data were suitable for analysis of IIO.

Tables 2 – 4 show the effect, sequentially, of removing items with the highest significant number of violations of IIO in the total baseline sample and the males and females separately.

In all three tables the value of H^T increases as items are removed. In the total sample 8 items are retained at the point where there were no further significant violations of IIO and for the male and females, respectively, 6 items and 5 items were retained. The male and female samples have 4 items showing IIO in common. Table 5 shows the Mokken scale extracted at baseline on the total sample ordered in terms of increasing item difficulty from top to bottom of the table. The items showing IIO in the total sample and in males and females is also shown and this reflects the data presented in Tables 2 – 4.

Table 6 shows the effect over the course of the study on a range of MSA parameters. Samples size generally decreases over the course of the study, although it increases from baseline to the first 6 months as a result of their being more missing data at baseline than at the first six-month assessment. Thereafter, attrition over the course of the study as participants die or are transferred to another nursing home leads to progressively decreasing sample size. Over the course of 5 years, and as a result of the reduced sample size, there is little effect on *Hs* or H^T and the main effects are that as the sample size decreases, up to a sample size of approximately 2000, all the 95% confidence intervals of *Hi* exclude the lower bound value 0.3. Also, at the highest samples sizes there are no violations of IIO but these become apparent at a sample size of approximately 7000 and are highest at the lowest sample size of approximately 500. Table 7 compares the mean item scores for males and females at baseline. For each ADL there is a statistically significant difference between males and females with women showing higher scores—and, thereby, greater dependence—for each item. In general, the effect sizes are quite small and the statistically significant difference between total scores persists when age is entered as a covariate. There was a significant increase in total ADL score (indicating greater disability) over the course of the study (Supplementary file 5).

4. DISCUSSION

This study aimed to investigate if there was a sustainable hierarchy of ADL items in a large database collected routinely in one region of Italy and adds to the growing body of knowledge about the use of IRT methods to study ADLs (Fieo et al., 2011). To date, this is the largest study available to our knowledge including older participants admitted to nursing homes and also longitudinally evaluated in terms of ADL performance in 11 items across 5 years. Previous studies have included limited sample sizes (McDonals and Ruhe, 2010) and have adopted multiple cross-sectional study design (Finlayson et al., 2005; Pascazio et al., 2009) analysing a limited number of ADLs or IADL and ADL items together (Pascazio et al., 2009). This is also the first study evaluating if there is a Mokken scale in this particular ADL dataset.

The results show that there is a hierarchy of ADL items which is sustained precisely across all the time points in the study. With the exclusion of the item on nutrition, the remaining items form a strong, reliable Mokken scale. The exclusion of the nutrition item was due to its insensitivity to the latent trait whereby it showed a floor effect and thus would also exaggerate the scaling properties (Meijer and Egberink, 21012). It is unclear whether this is due to there being no nutritional disturbance or that the item is very hard to assess. Moreover, the observers may also consider the declining food intake and body weight observed in the later stages of life to be part of the normal progression of physiological decline observed during aging (McDonald and Ruhe, 2010).

Sensitivity analysis demonstrated that females showed significantly greater disability than males and this is consistent with the observation that female gender is a predictor of nursing home entry (Caljouw et al., 2014). In the total sample the anchors for the hierarchy were urinary incontinence and feeding—both showing IIO— with feeding indicating the lowest level of difficulty. Urinary incontinence is well recognised as a predictive factor influencing nursing home admission in older individuals (Maxwell et al., 2013) and the general decline in ADLs reflects previous studies (Cohen-Mansfield et al., 1995; Kempen et al., 1995). However, it is worth noting that, while feeding represents the lowest difficulty in males, it does not in females and in both genders the use of stairs is the anchoring item at the highest level of difficulty. These observations may have clinical relevance, indicating that a common set of ADL items may not consistently assess decline in ability in males and females. However, this may be an effect of the reduced sample size relative to the total sample, when males and females are analysed separately. This effect is demonstrated by the analysis of Mokken scaling parameters in the 8 items showing IIO in the total sample across the whole study. As the sample size decreases across the waves of the study, violations of IIO appear and the 95% confidence intervals increase.

The clinical relevance of this study lies in the demonstration of the hierarchy of items in this database and that the hierarchy is sustained as disability increases over time. Therefore, the utility of this set of items as predictors of disability and as an indicator of the actual pattern of in decline of ADLs has been increased. Further work is required to investigate the possibility of gender sensitivity in this set of items. Also, there may be other independent variables such as age (for example, 'old' versus 'oldest old'), body mass index an cognitive status which may influence the declining pattern of ADL and these will be worth investigating in future. Further work is also needed to increase the predictive utility of these items by scaling alongside other less severe ADL disability indicators—for example as contained in the Townsend Functional Ability Index (Fieo et al., 2009)—to see if there are congruent indicators of the levels of disability demonstrated here in community dwelling older people. Extending the scale to lower levels of disability would increase its ability to predict, for example, the potential need for nursing home admission or earlier preventive interventions.

REFERENCES

Caljouw, M.A., Cools, H.J., Gussekloo, J. 2009. Natural course of care dependency in residents of long-term care facilities: prospective follow-up study. BMC Ger. 22, 14-67.

Cohen-Mansfield, j., Werner, P., Resiberg, B. 1995. Temporal order of cognitive ad functional loss in a nursing home population. J. Amer. Geriat. Soc. 43, 974-8.

Daltroy, L.H., Logigian, M., Iversen, M.D., Liang, M.H. 1992. Does musculoskeletal function deteriorate in a predictable sequence in the elderly? Arth. Care. Res. 5, 146-50.

Dunlop, D.D., Hughes, S.L., Manheim, L.M. 1997. Disability in activities of daily living: patterns of change and a hierarchy of disability. Am. J. Pub. Health. 87, 378-383.

Fieo, R.A., Austin, E.J., Starr, J.M., Deary, I.K. 2011. Calibrating ADL-IADL scales to improve measurement accuracy and to extend the disability construct into preclinical range: a systematic review BMC Ger. [serial online] 2011 August [cited 2015 Sept 1; 11, 42 [12 screens]. Available from : URL: http://www.biomedcentral.com/1471-2318/11/42.

Fieo, R., Watson, R., Deary, I.J., Starr, J.M. 2009. A revised activities of daily living/instrumental activities of daily living instrument increases interpretive power: theoretical application for functional tasks exercise Gerontology. 56, 482-90.

Finlayson, M.1., Mallinson, T., Barbosa, V.M. 2005. Activities of daily living (ADL) and instrumental activities of daily living (IADL) items were stable over time in a longitudinal study on aging. J. Clin. Epid. 58, 338-49.

Hebert. R, Bilodeau, C.A. 1988. The Functional Autonomy Measurement System (SMAF): description and validation of an instrument for the measurement of handicaps. Age. Ageing. 17, 293-302.

Hjaltadóttir, I., Hallberg, I.R., Ekwall, A.K., Nyberg, P. 2012. Health status and functional profile at admission of nursing home residents in Iceland over 11-year period. Int. J. Old. People. Nurs. 7, 177-87.

Gerrard, P. 2013. The hierarchy of the activity Daily Living in the Katz index in residents in skilled nursing facility. J. Geriart, Phys. Ther. 3, 87-91.

Katz, P.P. for the Association of Rheumatology Health Professionals Outcomes Measures Task Force.2003. Measures of adult general function. Arth. Care. Res. 49, S15-17.

Kruse, R.L., Petroski, G.F., Mehr, D.R., Banaszak-Holl, J., Intrator, O. 2013. Activity of daily living trajectories surrounding acute hospitalization of long-stay nursing home residents. J Amer. Ger. Soc. 61, 1909-1918.

Kempen, G.I.J.M., Myers, A.M., Powell, L.E. 1995. Hierarchical structure in ADL and IADL: analytical assumptions and applications for clinicians and researchers. J. Clin. Epid. 48, 1299-305.

Kuijpers, R.E., van der Ark, L.A., Croon, M.A. 2013. Standard errors and confidence intervals for scalability coefficients in Mokken scaling analysis using marginal models. Sociol. Meth. 43, 42-69.

Lawton, M.P., Brody, E.M. 1969. Assessment of older people: Self-maintaining and instrumental activities of daily living. Gerontologist. 9, 179-86.

Lazardis, E.N., Rudberg, M.A., Furner, S.E., Cassel, C.K. 1994. Do activities of daily living have a hierarchical structure? An analysis using the Longitudinal Study of Aging. J. Gerontol. 9, M47-M51.

Liu, W., Unick, J., Galik, E., Resnick, B. 2015. Barthel Index of activities of daily living Nurs. Res. 42, 88-99.

McDonald, R.B.1., Ruhe, R.C. 2010. Changes in food intake and its relationship to weight loss during advanced age. Interdisc. Top. Gerontol. 37, 51-63.

Mahoney, F.I., Barthel, D.W. 1965. Functional evaluation: The Barthel Index. A simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill. Mary St. Med. J. 14, 56-61.

Maxwell, C.J., Soo, A., Hogan, D.B., Wodchis, W.P., Gilbart, E., Amuah, J., Eliasziw, M., Hagen, B., Strain, .LA. 2013. Predictors of Nursing Home Placement from Assisted Living Settings in Canada. Can. J. Age. 4, 1-16.

Meijer, R.R., Egberink, I.J.L. 2012. Investigating invariant item ordering in personality and clinical scales: some empirical findings and a discussion. Educ. Psych. Meas. 72, 589-607.

Meijer, R.R., Sijtsma, K., Smid, N.G. 1990, Theoretical and empirical comparison of the Mokken and Rasch approach to ITR. App. Psychol. Meas. 14, 283-98.

Mokken, R.J., Lewis, C. 1982. A nonparametric approach to the analysis of dichotomous item responses. App. Psychol. Meas. 6, 417-30.

Palese, A., Menegazzi, J., Tullio, A., Zigotti, M., Hayter, M., Watson, R. 2016. Functional decline in residents living in nursing homes: a systematic review of the literatura. J. Amer. Med. Dir.Assoc. doi: 10.1016j.jamda.2016.04.002.

Pascazio, L., Morosini, P., Bembich, S., Nardone, I., Clarici, A., Barbina, L., Zuttion, R., Gigantesco, A. 2008. Description and validation of a geriatric multidimensional graphical instrument for promoting longitudinal evaluation. Arch. Gerontol. Ger. 48, 317-24.

Roskam, E.A., van den Wollenberg, A.L., Jansen, P.G.W. 1986. The Mokken scale: a critical discussion. App. Psychol. Meas. 10, 265-77.

Sijtsma, K., Meijer, R.R., van der Ark, A. 2011. Mokken scale analysis as time goes by: an update for scaling practitioners. Pers. Ind .Diff. 50, 31-37.

Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarsfeld, P.F., Star, S.A., Clausen, J.A. 1950. Measurement and prediction. Princeton: Princeton University Press.

The Medical Research Council Cognitive Functioning and Ageing Study (MRC CFAS). 1998. The description of activities of daily living in five centers in England and Wales. Gerontologist. 27, 605-13.

van der Ark, L.A. 2014. Package ‘mokken’. 2014 August [cited 2015 Sept 1]; [pdf] Available from: URL: https://cran.r-project.org/web/packages/mokken/mokken.pdf.

van Schuur, W.H. 2003. Mokken scale analysis: between the Guttman scale and parametric item response theory. Polit. Anal. 11, 139-63.

Watson, R. 1996. Mokken scaling procedure (MSP) applied to feeding difficulty in elderly people with dementia. Int. J. Nurs. Stud. 33, 385-93.

Watson, R., van der Ark, L.A., Lin, L-C., Fieo, R., Deary, I.J., Meijer, R.R. 2012. Item response theory: How Mokken scaling can be used in clinical practice J. Clin. Nurs. 21, 2736-46.

Watson, R., Thompson, D.R., Wang, W. 2015. Violations of local stochastic independence exaggerate scalability in Mokken scaling analysis of the Chinese Mandarin SF-36. Health. Qual. Life. Out. [serial online] 2014 October [cited 2015 Sept 1; 12, 149 [9 screens]. Available from: URL: http://www.hqlo.com/content/12/1/149.

Table 1

Activities of daily living included in the present study

Activity of daily living	Scoring
Self-feeding	0 (independent) – 5 (dependent)
Appetite disturbance/nutrition	0 (none) – 5 (refuse to eat)
Bathing	0 (independent) – 5 (dependent)
Bladder continence/urinary continence	0 (continent) – 6 (incontinent)
Faecal continence	0 ((continent) – 6 (incontinent)
Able to use toilet	0 (independent) – 5 (dependent)
Dressing	0 (independent) – 5 (dependent)
Self-care (cleaning face and arms, combing, brushing teeth, making up)	0 (independent) – 5 (dependent)
Stairs (20 steps)	0 (independent) – 5 (dependent)
Walking (50 metres)	0 (independent) – 5 (dependent)
Transfer from bed to armchair	0 (independent) – 5 (dependent)
Mobility in bed	0 (independent) – 5 (dependent)
Use of physical aids	0 (independent) – 5 (dependent)

Table 2

Number of significant violations of IIO and effect on Htrans (H^T) in the total sample with sequential item removal steps (baseline sample; N = 11,500)

Item	Label	Step 1	Step 2	Step 3	Step 4	Step 5
1	Feeding	0	1	1	0	0
2	Bath or shower	12	-	-	-	-
3	Bladder continence/urinary incontinence	5	3	3	2	0
4	Fecal continence	0	0	0	0	0
5	Able to use toilet	6	4	2	1	0
6	Dressing	5	7	-	-	-
7	Self-care	6	6	7	-	-
8	Stairs (20 steps)	4	3	3	3	-
9	Walking (50 meres)	3	3	2	0	0
10	Transfer from bed to armchair	9	6	3	2	0
11	Mobility in bed	0	0	0	0	0
12	Use of physical aids (wheelchair)	0	1	0	0	0
	H^T	.48	.50	.51	.55	.56

Table 3

Number of significant violations of IIO and effect on Htrans (H^T) in males with sequential item removal steps (baseline sample; N = 2,885)

Item	Label	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
1	Feeding	0	1	0	0	0	0	0
2	Bath or shower	12	-	-	-	-	-	-
3	Bladder continence/urinary incontinence	8	3	4	2	-	-	-
4	Fecal continence	0	0	0	0	2	1	0
5	Able to use toilet	9	6	3	0	0	0	0
6	Dressing	6	7	8	-	-	-	-
7	Self-care	7	8	-	-	-	-	-
8	Stairs (20 steps)	6	2	2	2	1	0	0
9	Walking (50 meres)	0	2	1	0	1	1	0
10	Transfer from bed to armchair	8	8	4	2	2	-	-
11	Mobility in bed	2	3	2	2	1	3	0
12	Use of physical aids (wheelchair)	2	1	2	2	1	3	-
	H^T	.43	.43	.46	.48	.50	.50	.65

Table 4

Number of significant violations of IIO and effect on Htrans (H^T) in females with sequential item removal steps (baseline sample; N = 8,615)

Item	Label	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8
1	Feeding	0	0	0	0	0	1	-	-
2	Bath or shower	12	-	-	-	-	-	-	-
3	Bladder continence/urinary incontinence	5	3	3	-	-	-	-	-
4	Fecal continence	0	0	0	0	0	0	1	-
5	Able to use toilet	5	3	2	2	1	0	0	0
6	Dressing	6	7	2	4	-	-	-	-
7	Self-care	4	5	1	3	4	-	-	-
8	Stairs (20 steps)	4	4	3	0	0	0	0	0
9	Walking (50 meres)	3	2	1	5	3	0	1	0
10	Transfer from bed to armchair	11	10	-	-	-	-	-	-
11	Mobility in bed	0	0	0	0	0	0	0	0
12	Use of physical aids (wheelchair)	0	0	0	0	0	1	0	0
	H^T	.50	.52	.53	.52	.52	.57	.55	.69

Table 5

Mokken scale of items retained at baseline

Item	Label	Mean score	*H* (SE)
3	Bladder continence/urinary incontinence	3.28	0.72 (0.005)*
8	Stairs (20 steps)	2.97	0.79 (0.004) †✝
2	Bath or shower	2.87	0.77 (0.004)
5	Able to use toilet	2.63	0.82 (0.003)* †✝
6	Dressing	2.58	0.82 (0.003)
10	Transfer from bed to armchair	2.44	0.80 (0.003)*
7	Self-care	2.33	0.76 (0.004)
9	Walking (50 meres)	2.18	0.78 (0.003)* †✝
4	Fecal continence	1.98	0.74 (0.004)* †
11	Mobility in bed	1.70	0.76 (0.004)* †✝
12	Use of physical aids (wheelchair)	1.46	0.77 (0.004)* ✝
1	Feeding	1.17	0.72 (0.005)* †

Hs = 0.78; H^T = 0.59; Rho = 0.96; * = items showing IIO in total scale; † = items showing IIO in males; ✝ = items showing IIO in females

Table 6

Effect of varying sample size on Mokken scaling parameters (8 items)

	1	2	3	4	5	6	7	8	9	10	11
n	11,500	11,678	9083	7259	5669	4210	2916	2090	1494	949	568
Hs (12 items)	.78	.76	.77	.77	.77	.78	.78	.78	.79	.80	.82
H_i with 95% CI<.3	0	0	0	0	0	0	0	0	1	1	0
H^T	.56	.57	.58	.58	.59	.59	.58	.59	.60	.61	.59
Significant violations of IIO	0	0	0	2	0	2	2	2	0	2	4

Table 7

Differences in activities of daily living mean scores between females (n=8615) and males (n= 2885)

	Mean score				
Activity of daily living	Women	Men	Difference (95% CI)	p*	d
Self-feeding	1.23	1.01	.21 (.15-.27)	<.001	.15
Bathing	2.93	2.72	.21 (.16-.26)	<.001	.19
Bladder continence/urinary continence	3.43	2.84	.59 (.50-.69)	<.001	.28
Faecal continence	2.06	1.73	.33 (.25-.41)	<.001	.19
Able to use toilet	2.74	2.29	.45 (.37-.52)	<.001	.27
Dressing	2.64	2.38	.27 (.21-.33)	<.001	.19
Self-care	2.37	2.20	.17 (.10-.23)	<.001	.11
Stairs (20 steps)	3.10	2.59	.51 (.44-.57)	<.001	.33
Walking (50 metres)	2.29	1.86	.42 (.35-.49)	<.001	.25
Transfer from bed to armchair	2.57	2.04	.53 (.45-.61)	<.001	.29
Mobility in bed	1.82	1.35	.47 (.40-.54)	<.001	.27
Use of physical aids	1.56	1.27	.29 (.23-.35)	<.001	.20
Total score	28.20	25.28	4.49 (3.83-5.16)	<.001**	.18

* = t-test; ** = ANCOVA with age as the covariate; d = Cohen's effect size