Psychometrics of the observational scales of the Utrecht Scale for Evaluation of Rehabilitation (USER): Content and structural validity, internal consistency and reliability: A study on psychometric properties of the USER.

Jansen, M.; Doornebosch, A.J.; Waal, M.W.M. de; Wattel, E.M.; Visser, D.; Spek, B.; Smit, E.B.

Archives of Gerontology and Geriatrics 97 (2021) 104509

Contents lists available at ScienceDirect

Archives of Gerontology and Geriatrics

journal homepage: www.elsevier.com/locate/archger

Psychometrics of the observational scales of the Utrecht Scale for Evaluation of Rehabilitation (USER): Content and structural validity, internal consistency and reliability

Michael Jansen [a,f,*], Arno J Doornebosch [b], Margot WM de Waal [b], Elizabeth M Wattel [c], Dennis Visser [c], Bea Spek [d], Ewout B Smit [c,e]

[a] Faculty of health, Physiotherapy, University of applied sciences Leiden, Leiden, The Netherlands
[b] Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, The Netherland
[c] Department of Medicine for Older People Amsterdam University Medical Centres - VU Amsterdam, Amsterdam, The Netherlands
[d] Department of Epidemiology and Data Science, Amsterdam University Medical Centres, University of Amsterdam, Amsterdam, The Netherlands
[e] Vivium Zorggroep Naarden, The Netherlands
[f] Woon Zorgcentra Haaglanden (WZH), The Hague, The Netherlands

ARTICLE INFO

Keywords:
Geriatric rehabilitation
Post-acute care
Clinimetrics
Physical functioning
Cognitive functioning

ABSTRACT

Introduction: : Establish content and structural validity, internal consistency, inter-rater reliability, and measurement error of the physical and cognitive scales of the Utrecht Scale for Evaluation clinical Rehabilitation (USER) in geriatric rehabilitation.

Material and methods: : First, an expert consensus-meeting (N=7) was organised for content validity wherein scale content validity index (CVI) was measured. Second, in a sample of geriatric rehabilitation patient structural validity (N=616) was assessed by confirmatory factor analyses for exploring unidimensionality. Cut-off criteria were: Root Mean Square Error of Approximation (RMSEA) $\leq$0.08; Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) $\geq$0.95. Local independence (residual correlation<0.20) and monotonicity (H_i-coefficient $\geq$0.30 and H_s-coefficient $\geq$0.50) were also calculated. Cronbach alphas were calculated for internal consistency. Alpha's > 0.7 was considered adequate.

Third, two nurses independently administered the USER to 37 patients. Intraclass-correlation coefficients (ICC) were calculated for inter-rater reliability (IRR), standard error of measurement (SEM) and Smallest Detectable Change (SDC).

Results: : The CVI for physical functioning was moderate (0.73) and excellent for cognitive functioning (0.97). Structural validity physical scale was acceptable (CFI;0.95, TLI;0.93, RMSEA;0.07, ECV;0.78, OmegaH;0.87; Monotonicity;(H_i;0.52-0.75 and H_s;0.63)). Cognitive scale was good (CFI;0.98, TLI;0.96, RMSEA;0.05, ECV;0.66 and OmegaH;0.90. Monotonicity;(H_i;0.30 –0.70 and H_s;0.61)). Cronbach's alpha were high: physical scale;0.92 and cognitive scale;0.94. Reliability physical scale ICC;0.94, SEM;5 and SDC;14 and cognitive scale ICC;0.88, SEM;5 and SDC;13.

Conclusion: : The observational scales of the USER have shown sufficient content and structural validity, internal consistency, and interrater reliability for measuring physical and cognitive function in geriatric rehabilitation.

Trial registration: : N/A

1. Introduction

Geriatric Rehabilitation (GR) is a multidimensional approach of diagnostic and therapeutic interventions, which focuses on optimizing functional capacity, promote activity and preserve functional reserve and social participation in older people with disabling impairments (Grund et al., 2020, Bachmann et al., 2010). These patients are characterized by increased vulnerability, multimorbidity and decreased trainability and learnability (Grund et al., 2020, Bachmann et al., 2010). The latter stresses cognitive function as being an important factor in

* Corresponding author.
E-mail addresses: jansen.mic@hsleiden.nl (M. Jansen), eb.smit@amsterdamumc.nl (E.B. Smit).

https://doi.org/10.1016/j.archger.2021.104509
Received 16 June 2021; Received in revised form 21 August 2021; Accepted 24 August 2021
Available online 26 August 2021

geriatric rehabilitation. Also, lower cognitive function is negatively associated with both restoration of physical function and home discharge (Bachmann et al., 2010, Hartley et al., 2017, Everink et al., 2016). It is thus essential for evaluation of rehabilitation to measure both physical and cognitive function as important predictors for successful rehabilitation. In addition, both physical and cognitive function are an important components of the comprehensive geriatric assessment (CGA), which itself might be considered essential for geriatric rehabilitation.

When measuring physical and cognitive function it is important to use an instrument with adequate measurement properties (De Vet et al., 2011, Mokkink et al., 2010). A review showed that there is large variability in measuring and defining a geriatric patients general functional status and there is no conclusive evidence for psychometric superiority of one instrument (Buurman et al., 2011). In geriatric rehabilitation, the Barthel index (BI) and Katz index are probably the most commonly used instruments for assessing physical functioning of basic activities of daily live (ADL) (Bouwstra et al., 2019, Wallace & Shelkey, 2008). There are however some disadvantages to the use of the BI and Katz index, as they only measure basic ADL and not the whole range of physical functioning and do not measure cognitive functioning. Furthermore, the BI showed significant ceiling effects even with geriatric rehabilitation patients (Bouwstra et al., 2019).

A promising multidimensional instrument is the Utrecht Scale for Evaluation of clinical Rehabilitation (USER). The USER was developed in adult rehabilitation as a generic multidomain instrument or minicoreset to measure physical and cognitive function and self-reported pain, mood and fatigue. The developers considered the content and psychometric properties of existing instruments insufficient for the use in adult rehabilitation and therefore developed a new instrument (Post et al., 2009).

The primary development and validation study concluded that the USER is a reliable, valid and responsive instrument in adult rehabilitation patients (Post et al., 2009). However, the USER was never validated for geriatric rehabilitation patients, who exhibit more cognitive problems, multimorbidity and polypharmacy (Grund et al., 2020, Bachmann et al., 2010). Because of these differences it is unclear whether the psychometric properties of the USER are also sufficient for these patients.

In conclusion, currently there is insufficient evidence on the psychometric properties of the observational scales physical and cognitive function of the USER for use in geriatric rehabilitation (De Vet et al., 2011, Mokkink et al., 2010).

The objectives of this study are to evaluate the content and structural validity, internal consistency, inter-rater reliability, and measurement error of the physical and cognitive function scale of the USER in inpatient geriatric rehabilitation.

2. Material and methods

The USER is a multidimensional measurement instrument containing both observational and patient-reported scales. The observational scales measure physical function, divided in two scales, mobility and self-care, both containing seven items and cognitive function, containing ten items. The ordinal response options range from zero to five points with a higher score representing higher function (Wallace & Shelkey, 2008). The patient-reported scales, which measure pain, fatigue and mood, are not the scope of the current study, as we are mainly interested in cognitive and physical function and also require different testing procedures.

The current study, which was conducted according to COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (Mokkink et al., 2010) guidelines consisted of three parts: 1) a consensus meeting to establish content validity and two cross sectional studies to establish 2) structural validity and internal consistency, and 3) reliability, including measurement error.

2.1. Study 1: Content validity

2.1.1. Design and study population

We planned for a consensus meeting with a minimum of seven experienced professional experts in geriatric rehabilitation (Mokkink et al., 2010). Experts from two academic networks for the elderly care were invited to participate in the study in case they participate in a multidisciplinary geriatric rehabilitation team with at least two years working experience in GR. Experience with the USER was not a specific requirement. The meeting was moderated by the coordinating researcher (EBS) and assisted by two independent observers (AJD and MJ). The experts were either geriatric rehabilitation physicians (assistants), nurses, occupational therapists, or physiotherapists. The meeting started by explaining the purpose and process of the consensus meeting, followed by stating the definitions of the constructs physical and cognitive function . First, participants interacted and discussed to what degree all items are relevant for the scale and target population and whether instructions and response options per item were understood. We finished this phase if full consensus was reached. Next, participants were asked if items were missing. Participants independently rated all individual items on relevance after the meeting based on a three point scale: ‘essential’, ‘useful but not essential’, and ‘not necessary’. From that an item content validity index (CVR) was calculated, which can range from -1 to +1. Finally, a scale content validity index (CVI) was calculated on relevance (Lawshe, 1975).

2.1.2. Statistical analysis

A critical value CVR 1 was considered sufficient for relevance of items and a CVI $\geq$0.9 was considered excellent content validity (Lawshe, 1975). A CVI $\geq$0.8$\leq$0.9 was considered good content validity and $\geq$0.7$\leq$0.8 moderate. The formula used for calculating the CVR was;

$$CVR = \frac{Ne - N/2}{N/2}$$

(N_e= the number of experts that rated the item ‘essential’, N= number of experts).

The CVI was calculated as the mean of all items CVR.

2.2. Study 2: Structural validity and internal consistency

2.2.1. Design and study population

We hypothesized that the USER’s observational scales form two independent unidimensional scales (i.e. physical function combining the two scales for mobility and self-care to one subscale and cognitive function). We used routine care data in a cross-sectional design. Study participants were admitted to geriatric rehabilitation. Routinely collected data from 616 newly admitted patients administered in 2019-2020, within the age limits of 60-90 years, from a central university network database was used for this part of the study.

2.2.2. Statistical analysis

We assessed structural validity by evaluating unidimensionality, monotonicity and local independence. Unidimensionality of both scales was tested with two independent confirmatory factor analyses (CFA). Cut-off criteria for unidimensionality were a Root Mean Square Error of Approximation (RMSEA) $\leq$0.08, a Comparative Fit Index (CFI) and a Tucker Lewis Index (TLI) $\geq$0.95 (Hu & Bentler, 1999). If a one factor first order model for physical function failed to meet the cut-off criteria, a nested correlated factors model for physical function containing two subscales mobility and self-care and a hierarchical model would be assessed. If a one factor model for cognitive function failed to meet the cut-off criteria for structural validity, a nested correlated factors model containing three subscales communication, cognition and behaviour and a hierarchical model would be assessed.

If criteria for strict unidimensionality were not met, essential

unidimensionality (EUD) was assessed for both scales based on a bi-factor model (Reise et al., 2013). EUD is the case when a substantial part, but not all, of the variance in the data is explained by a single large general factor. Cut-off criteria for EUD were an explained common variance (ECV) >0.6 and an OmegaH >0.8.

Monotonicity for measuring scalability was examined with the Mokken package in R version 3.0.2. Criteria for monotonicity acceptability were item scalability coefficients $H_i \geq 0.30$ and scalability coefficient of the scale $H_s \geq 0.50$ (Mokken, 1971).

Finally, local independence, which is the assumption that, conditional on the latent variable(s), item responses are independent was examined by analysing residual correlations. Cut-off criteria for violations were residual correlations between items >0.20 (Edwards et al., 2018).

We applied a rule of thumb with seven to ten participants per item and ≥200 for the total sample size for the confirmatory factor analysis (Mokkink et al., 2010, Kyriazos, 2018). A weighted least square mean and variance estimator method was used (WLSMV) (Li, 2016). All items were examined for non-normal distribution both visual and formal with the multivariate normality (MVN) version 5.8 package in R version 3.6.3.

The Cronbach's alpha of each scale was calculated to determine internal consistency. Cronbach alpha scores >0.7 were considered adequate. To determine whether each individual item correlates well with the scale overall, an item total correlation was analysed. Values <0.3 are indicative of insufficient item-total correlation (Mokkink et al., 2010, Tavakol & Dennick, 2011).

2.3. Study 3: Reliability

2.3.1. Design and study population

The reliability of the USER was determined by evaluating its interrater reliability and measurement error. We used a cross sectional design with a convenience sample of geriatric rehabilitation patients from two wards in one geriatric rehabilitation facility. A single measure was independently done by two trained and experienced nurses on each patient. We anticipated an Intraclass correlation coefficient (ICC) of 0.8 and a 95% confidence interval of 0.2 for calculating Interrater Reliability (IRR). The required sample size for two repeated measurements was a minimum of 14 (De Vet et al., 2011, Mokkink et al., 2010, Tavakol & Dennick, 2011, Giraudeau & Mary, 2001).

2.3.2. Statistical analysis

We calculated intraclass correlation coefficient (ICC), standard error of measurement (SEM) and smallest detectable change (SDC) for both scales separate. For the calculation of the ICC we used a two-way random, singles measure absolute agreement definition. The following formulas were used (De Vet et al., 2011, Mokkink et al., 2010);

$$ICCagreement = \left(\sigma_p^2\right) \Big/ \left(\sigma_p^2 + \sigma_o^2 + \sigma_{residual}^2\right)$$

$$SEMagreement = \sqrt{\left(\sigma_o^2 + \sigma_{residual}^2\right)}$$

$$SDC = 1,96\ x\ SEMagreement\ x\ \sqrt{2}$$

(σ_p^2= person variance, σ_o^2= observer variance and $\sigma_{residual}^2$= residual variance error)

2.3.3. Handling of data

Data for measuring structural validity and interrater reliability was primarily processed in SPSS (Armonk, NY: IBM Corp.) version 26. The interrater reliability, structural validity and internal consistency were analysed using R version 3.6.3 (R foundation for statistical computing, Vienna, Austria. Packages irr version 0.84.1 and Lavaan version 0.6-6).

2.4. Ethical aspects

The study was approved by a Medical Ethics Review board of the Amsterdam University Medical Centres, location VUmc (FWA00017598).

3. Results

3.1. Content validity

The consensus meeting consisted of seven experts: two geriatric rehabilitation physicians (assistants), two nurses, one occupational therapist, and two physiotherapists, all with at least two years' experience in GR. Full consensus was reached during the consensus meeting and the experts concluded the USER could be a valid measure of functional status in geriatric rehabilitation In addition, the experts agreed that the instructions could be written clearer for both scales.

Table 1 shows that the scale content validity (CVI) for the physical scale was moderate (0.73). Four items of the physical scale did not meet the minimum item content validity (CVR) criterion of 0.78. These were "walking longer distance" (0.43), "riding a wheelchair" (0.71) and the items on "bowel and bladder incontinence" (both -0.43). In that line, the experts stated that frequency of incontinence was not considered by them a part of the construct physical function. According to the experts scoring the level of independence of handling incontinence instead would be more logical in measuring a construct of physical functioning. The expert panel suggested that bowel and bladder incontinence could be marked as obsolete items when combined with the item toileting. The item-scoring "walking outdoors" was found inappropriate for the target population because the distance of thousand meters was considered too long for geriatric rehabilitation patients. Nonetheless all agreed that this item should remain in the physical scale but scoring options and description could be altered. The scoring options of the item "eating and drinking" was considered difficult to score, in particular to distinguish between "help" and "with difficulty". Experts agreed that a separate item for scoring a person's ability shifting/moving in bed could be added to the physical scale.

The cognitive scale had an excellent total scale content validity (0.97), only the item "social behaviour" did not meet the minimum CVR criterium (0.71). Experts expressed having trouble to score non-verbal understanding for the item "understanding". Experts agreed that a new item scoring a person's ability to learn new skills could be added.

Table 1
Item content validity index and scale content validity index of the USER physical and cognitive scales

Item and scale content validity index					
Physical scale	CVR	CVI	Cognitive scale	CVR	CVI
item		0.73	item		0.97
Sit	1		Express oneself	1	
Stand	1		Understanding	1	
Transfers	1		Visual perception	1	
Walking indoors	1		Orientation in time and place	1	
Walking longer distance	0.43		Attention	1	
Walking stairs	1		Memory	1	
Wheelchair driving	0.71		Task performance	1	
Eating and drinking	1		Initiative	1	
Personal care	1		Behaviour regulation	1	
Showering/ bathing	1		Social behaviour	0.71	
Dress and undress	1				
Toileting	1				
Incontinence bladder	-0.43				
Incontinence bowel	-0.43				

3.2. Structural validity and internal consistency

Table 2 shows the descriptive data for the sample used for the analysis of the structural validity and reliability. The mean age in the sample was 78 (SD8,3; 60-90). There was no missing data and the distribution of the data for both scales was moderately non-normal.

For the physical scale, a second order hierarchical model, with a general factor and two with the general factor correlating underlying sub-constructs had the best fit for structural validity. Fit measures were; CFI 0.95, TLI 0.93 and RMSEA 0.07 (90%CI;0.06-0.08, $p < 0.001$). Item factor loadings were all above 0.4 (0.51-0.92) with exception for item six (walking stairs 0.33). Removal of item six did not significantly improve fit indices. The ECV for the physical model was 0.78 with an OmegaH global of 0.87. There were no significant violations for monotonicity (H_i 0.52-0.75 and H_s 0.63) and there were violations on local independence between the items "walking longer distance" with "stairs climbing" (0.23) and between the "incontinence" items (0.26).

For the cognitive scale, a second order model had the best fit for structural validity. Fit measures were; CFI 0.98, TLI 0.96 and RSMEA 0.05 (90%CI;0.03-0.07, p=0.05). Item factor loadings were all above 0.4 (0.82-0.91) with exception for item seventeen (visual perception 0.33). Removal of item seventeen did not significantly improve fit indices. The ECV for the cognitive scale was 0.66 with an OmegaH global of 0.90 (table 3). There were no significant violations for monotonicity (H_i0.30–0.70 and H_s0.61) and no violations on local independence.

The Cronbach's alpha of the physical scale was 0.92 (95%CI;0.91-0.93) and for the cognitive one factor scale 0.94 (95%CI;0.93-0.95). Corrected item-total correlations were all above 0.4 with exception of "walking stairs" (0.36) for the physical scale and "visual perception" (0.38) for the cognitive scale. However, removal of these items did not change the internal consistency of the scales.

3.3. Inter-rater reliability and measurement error

Table 1 shows that a sample of 37 patients was used for this analysis, 65% were women and 35% were men. The mean age was 74 (SD13.4; 36-91). Data for the reliability study was normally distributed. The physical scale ICC was 0.94 (95%CI; 0.88-0.97), SEM was 5 and the SDC was 14.The cognitive scale ICC was 0.88 (95%CI;0.78-0.94), SEM was 5 and the SDC was 13.

4. Discussion

We tested the psychometric properties of the USER observational scales in inpatient geriatric rehabilitation and found that they have sufficient content and structural validity, internal consistency, and reliability for measuring physical and cognitive function.

Both scales had sufficient content validity. Still, several items had suboptimal fit with the intended construct for this patient group according to the experts. This was in particular the case for the physical scale. Experts suggested that instead of scoring the frequency of incontinence, the USER should score one's level of independence handling incontinence. This seems logical, two important health related quality of life models in geriatric rehabilitation, the International Classification of functioning, Disability and Health and Wilson and Cleary, classify continence itself as a biological or body functions, and not so much as a functional status or activity variable (World Health Organization International Classification of Functioning, Disability & Health 2001, Wilson & Cleary, 1995). For the item "walking outdoors" experts suggested altering the distance of 1000 meters to a shorter distance. It is important to note that a new version of the USER (1.5) has recently been released in which the item scoring has been changed to a distance of 250 meters.

We were not able to show sufficient fit for strict unidimensionality for the physical scale, still the criteria for essential unidimensionality were met. In other words, the degree of multidimensionality is not severe enough to disqualify the unidimensional nature of the instrument (Kyriazos, 2018). This is evidence that both scales of the USER can be considered unidimensional, which is important for use of total scale scores (Edwards et al., 2018). The structural validity of other instruments for geriatric rehabilitation patients has been tested. One study reported partially acceptable fit for structural validity of the BI in geriatric rehabilitation (Bouwstra et al., 2019). Several other studies have reported on psychometric properties of the Functional Independence Measure (FIM) for the use in geriatric patients (Dallmeijer et al., 2005, Ravaud et al., 1999, Nayar et al., 2016, Gunn et al., 2018). There is at best conflicting evidence for the structural validity of the FIM and to our knowledge there is no study exploring its content validity (Ladislav & Beat, 2018, Glenny & Stolee, 2009). In addition, one older study directly compared the validity of the BI and FIM and concluded that both instruments had similar psychometric properties (Hsueh et al., 2002).

Table 2
Patient characteristics of study samples 2 and 3, the interrater reliability and unidimensionality/internal consistency study

	Interrater Reliability study (n=37)	structural validity and internal consistency study (n=616)
Gender, n (%),		
Male	13 (35%)	252 (41%)
Female	24 (65%)	364 (59%)
Age, mean (SD; range)	74 (13.4; 36-91)	78 (8.3; 60-90)
Admission main diagnosis, n (%)		
Neurological	3 (8.1%)	NA
Orthopaedic (elective)	16 (43.2%)	NA
Oncology	7 (18.9%)	NA
Amputation	2 (5.4%)	NA
Internal	9 (24.3%)	NA

NA; Not Assessed, SD; standard deviation.

Table 3
Content of the USER with item total correlations, factor loadings and explained variance

Item	Content item	Corrected Item total correlation	Factor loading	Explained variance
	Physical scale			
1	Sitting	0.64	0.62	0.38
2	Standing	0.86	0.91	0.83
3	Transferring	0.86	0.92	0.85
4	Walking indoors	0.80	0.84	0.71
5	Walking outdoors	0.66	0.68	0.46
6	Walking stairs	0.36	0.33	0.11
7	Wheelchair riding	0.58	0.54	0.29
8	Eating and drinking	0.55	0.51	0.26
9	Grooming	0.71	0.73	0.53
10	Washing and showering	0.71	0.76	0.58
11	Dress/undress	0.79	0.84	0.71
12	Toileting	0.74	0.78	0.61
13	Incontinence bladder	0.66	0.62	0.38
14	Incontinence bowel	0.65	0.60	0.36
	Cognitive scale			
15	Express themselves	0.79	0.82	0.67
16	Understanding	0.83	0.86	0.74
17	Visual perception	0.38	0.33	0.11
18	Orientation in place and time	0.85	0.84	0.71
19	Attention and concentration	0.87	0.88	0.77
20	Memory	0.85	0.88	0.77
21	Task performance	0.88	0.85	0.72
22	Initiative	0.81	0.84	0.71
23	Behavioural regulation	0.81	0.82	0.67
24	Social behaviour	0.72	0.81	0.66

Although this study was not set up to directly compare different measurements a preliminary conclusion might be that validity of the USER is at least as good as the FIM and BI with the advantage of content validity for geriatric patients.

Both the physical and the cognitive scales had excellent internal consistency with item six "walking stairs" and item seventeen "visual perception" having a lower than 0.4 item-total correlation, suggesting these items do not correlate well with the other items. However, we think that removing these items would not give relevant changes as the low item-total correlations of these items on the total internal consistency is negligible and the items were found relevant by experts. The internal consistencies found in this study are comparable to those found for the BI and the FIM (Bouwstra et al., 2019, Ladislav & Beat, 2018).

The interrater reliability for both scales was excellent, which is in accordance with the original study of the USER (Post et al., 2009). Also, reliability outcomes for the USER from this study are comparable to those reported for the BI and the FIM in GR (Bouwstra et al., 2019, Glenny & Stolee, 2009). This means that USERs administered by different nurses show excellent agreement and can be used interchangeably, although we acknowledge the fact that training is important. In addition to the original study we calculated the measurement error, which was acceptable for both scales.

There are some limitations to our study. First, we did not have detailed information on the participating patients in study 2 due to general data protection regulation rules. Still, routine care data from a geriatric rehabilitation population were used without exclusion criteria. Therefore we expect that our patients are representative for geriatric rehabilitation. Second, we used admission data only for the analysis of structural validity, which could potentially include data with significant floor effects, which could in turn cause nuisance (i.e. method effects) in the analysis (Brown, 2015). And last, since there is no gold standard measure we did not extensively but only marginally compare the USER psychometric outcomes to other frequently used instruments in rehabilitation like the BI or the FIM and did not analyse correlation with other frequently used instruments like the BI and the FIM. However, the original development study of the USER reports strong correlations between the physical and cognitive scores of the USER, the BI and the FIM (Post et al., 2009).

A strength of this study is that we used routinely collected data for the analyses from a central database where data is collected from multiple GR clinics which might improve the external validity of our results. Another strength is that we tested the content validity of the USER by both qualitative (consensus meeting) and quantitative methods (calculation CVR), resulting potentially in a more thorough analysis from experts view on the USER than if only a quantitative analysis was done for example by means of a survey.

For future research we recommend establishing interpretability and responsiveness. The minimal important change (MIC) should be determined so that clinicians can give meaningful interpretation to change scores. Also, IRT scores should be established, so clinicians can efficiently define a patient's ability on physical and cognitive function (De Vet et al., 2011, Mokkink et al., 2010).

5. Conclusions

The observational scales of the USER have shown sufficient content and structural validity, internal consistency, and interrater reliability for measuring physical and cognitive function in geriatric rehabilitation. These findings strengthen dissemination and use of the observational scales of the USER within geriatric rehabilitation settings.

Brief summary

A study on psychometric properties of the USER.

Ethics approval and consent to participate

The study was approved by a Medical Ethics Review board of the Amsterdam University Medical Centres, location VUmc (FWA00017598).

Consent for publication

Routine care data from a central university network database was used with data collected with informed consent and an opt out option for patients.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on a reasonable request.

Funding

This study was supported by Verenso, the Dutch Association of Elderly Care Physicians (Verenso Grant 2017). Verenso had no active role in collection, analysis and interpretation of the data and in the writing of the manuscript.

Declaration of Competing Interests

All authors have no financial or non-financial conflicts of interest.

Acknowledgements

The authors would like to thank the experts participating in the consensus meeting: Annemiek Exterkate, Kitty van klink, Evelien Jansen, Danny van den Ende, Christa van Hest, Magdy Micheil and Renske Lelieveld.

References

Bachmann, S., Finger, C., Huss, A., Egger, M., et al. (2010). Inpatient rehabilitation specifically designed for geriatric patients: systematic review and meta-analysis of randomised controlled trials. *Bmj (Clinical Research Ed.), 340*(apr20 2). c1718-c1718.

Bouwstra, H., Smit, E. B., Wattel, E. M., van der Wouden, J. C., et al. (2019). Measurement Properties of the Barthel Index in Geriatric Rehabilitation. *JAMDA, 20* (4), 420-425.e1.

Brown, T. A. (2015). *Methodology in the social sciences. Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Buurman, B., van Munster, B., Korevaar, J., de Haan, R., & de Rooij, S. (2011). Variability in measuring (instrumental) activities of daily living functioning and functional decline in hospitalized older medical patients: a systematic review. *Journal of Clinical Epidemic, 64*(6), 619–627.

Dallmeijer, A., Dekker, J., Roorda, L., Knol, D., et al. (2005). Differential item functioning of the functional independence measure in higher performing neurological patients. *Journal of Rehabilitation Medicine, 37*(6), 346–352.

De Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide* (pp. 315–327). Cambridge: Cambridge University Press.

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departs from local independence in item response theory models. *Psychological Methods, 23* (1), 138–149.

Everink, I., van Haastregt, J., van Hoof, S., Schols, J, & Kempen, G. (2016). Factors influencing home discharge after inpatient rehabilitation of older patients: a systematic review. *BMC Geriatrics, 16*, 5.

Giraudeau, B., & Mary, J. (2001). Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Statistics in Medicine, 20* (21), 3205–3214.

Glenny, C., & Stolee, P. (2009). Comparing the functional independence measure & the interRAI/MDS for use in the functional assessment of older adults: a review of the literature. *BMC Geriatrics, 9*, 52.

Grund, S., Gordon, A. L., van Balen, R., Bachmann, S., et al. (2020). European consensus on core principles and future priorities for geriatric rehabilitation: consensus statement. *European Geriatric Medicine, 11*(2), 233–238.

Gunn, S., Burgess, G. H., & Maltby, J. (2018). A Factor Analysis of Functional Independence and Functional Assessment Mearsure scores Among Focal en Diffuse Brain Injury Patients: The Importance of Bifactor Models. *Archives of Physical Medicine and Rehabilitation, 99*(9), 1805–1810.

Hartley, P., Gibbins, N., Saunders, A., Alexander, K., et al. (2017). The association between cognitive impairment and functional outcome in hospitalised older patients: a systematic review and meta-analysis. *Age and Ageing, 1, 46*(4), 559–567.
Hsueh, I., Lin, J., Jeng, J., & Hsieh, C. (2002). Comparison of the psychometric characteristics of the functional independence measure, 5 item Barthel index, and 10 item Barthel index in patients with stroke. *Journal of Neurology, Neurosurgery, and Psychiatry, 73*(2), 188–190.
Hu, L., & Bentler, P. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structured Equation Model, 6*(1), 1–55.
Kyriazos, T. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology, 9*(8), 2207–2230.
Ladislav, V., & Beat, S. (2018). The dimensionality of functional independence measure (FIM). *International Physical Medicine and Rehabilitation Journal, 3*(5), 412–422.
Lawshe, C. H. (1975). A quantative approach to content validity. *Persistent Psychology, 28*, 563–575.
Li, C. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behaviour Research, 48*, 936–949.
Mokken, R. J. (1971). *A theory and procedure of scale analysis: with applications in political research*. De Gruyter: The Hague.
Mokkink, L., Terwee, C. B., Patrick, D., Alonso, J., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research, 19*(4), 539–549.
Nayar, M., Vanderstay, R., Siegert, R., & Turner-Stokes, L. (2016). The UK Functional Assessment Measure (UK FIM+FAM): Psychometric Evaluation in Patients Undergoing Specialist Rehabilitation following a Stroke from the National UK Clinical Dataset. *Plos One, 11*(1), Article e0147288.
Post, M. W. M., van der Port, I. G. L., Kap, B., & Berdenis van Berlekom, S. H. (2009). Development and validation of the Utrecht Scale for Evaluation of Clinical Rehabilitation (USER). *Clinical Rehabilitation, 23*(10), 909–917.
Ravaud, J. F., Delcey, M., & Yelnik, A. (1999). Construct validity of the functional independence measure (FIM): questioning the unidimensionality of the scale and the "value" of FIM scores. *Scandinavian Journal of Rehabilitation Medicine, 31*(1), 31–41.
Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*, 5–26.
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55.
Wallace, M., & Shelkey, M. (2008). Monitoring functional status in hospitalized older adults. *American Journal of Nursing, 108*(4), 64–71.
Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life: conceptual model of patient outcomes. *Journal of the American Medical Association, 273*, 59–65.
World Health Organization International Classification of Functioning, Disability and Health. World Health Organization; Geneva Switzerland: 2001.