



## COMPARISON OF THE DE MORTON MOBILITY INDEX AND HIERARCHICAL ASSESSMENT OF BALANCE AND MOBILITY IN OLDER ACUTE MEDICAL PATIENTS

Tobias BRAUN, MSc<sup>1</sup>, Christian GRÜNEBERG, PhD<sup>1</sup>, Anna COPPERS, BSc<sup>1</sup>, Linda TOFAUTE, BSc<sup>2</sup> and Christian THIEL, PhD<sup>1</sup>  
From the <sup>1</sup>Department of Applied Health Sciences, Physiotherapy Program, Hochschule für Gesundheit, Bochum, and <sup>2</sup>Catholic Clinic Bochum, Ruhr-University Bochum, Marien-Hospital Wattenscheid, Bochum, Germany

**Objective:** To compare the measurement properties of the de Morton Mobility Index (DEMMI) and the Hierarchical Assessment of Balance and Mobility (HABAM) in an older acute medical inpatient population.

**Design:** Cross-sectional.

**Subjects:** Older acute medical inpatients.

**Methods:** The DEMMI, HABAM and further assessments were performed after hospital admission. Construct validity was assessed by testing 13 hypotheses on convergent and known-groups validity. Test–retest reliability and minimal detectable change were estimated based on a re-assessment of unchanged patients. Floor and ceiling effects were used to indicate adequacy of scale width.

**Results:** For both the DEMMI and HABAM, 11 (85%) hypotheses regarding construct validity were confirmed ( $n=158$ ). Both scales showed strong correlations with other multi-component mobility scales (Spearman's rho 0.75–0.92). Neither floor nor ceiling effects were evident. The intraclass correlation coefficient was 0.98 (95% confidence interval (95% CI) 0.96–0.99) for the DEMMI and 0.99 (95% CI 0.99–0.99) for the HABAM, respectively ( $n=30$ ). The minimal detectable change with 90% confidence was 6 points on the 100-point DEMMI scale and 1 point on the 26-point HABAM scale.

**Conclusion:** The DEMMI and the HABAM appear to be suitable for measuring mobility in older acute medical patients.

**Key words:** frail elderly; mobility limitation; geriatric assessment; psychometrics.

Accepted Dec 18, 2017; E-pub ahead of print Jan 31, 2018

J Rehabil Med 2018; 50: 292–301

Correspondence address: Tobias Braun, Hochschule für Gesundheit, Department of Applied Health Sciences, Physiotherapy Program, Gesundheitscampus 6–8, DE-44801 Bochum, Germany. E-mail: tobias.braun@hs-gesundheit.de

Mobility limitations are frequent in hospitalized frail older individuals, with approximately 75% of these individuals being unable to walk or ambulating with a very low gait speed of  $\leq 0.6$  m/s (1, 2). These mobility limitations and the ensuing physical inactivity can increase the risk of mortality and functional decline in the hospital and after discharge (1, 3). Feasible, valid and reliable instruments to measure mobility are a

prerequisite to monitor rehabilitative interventions and to accurately track recovery from illness.

The International Classification of Functioning, Disability and Health (ICF) (4) defines “mobility” as “moving by changing body position or location or by transferring from one place to another, by carrying, moving or manipulating objects, by walking, running or climbing, and by using various forms of transportation”. Two systematic reviews (5, 6) have identified significant limitations of existing measurement tools of older patient's mobility, mainly due to large floor effects (Timed Up and Go test (TUG), gait speed) as many older patients are not able to stand and walk. The Hierarchical Assessment of Balance and Mobility (HABAM) (7–9) has been reported to be the instrument with the most desirable measurement properties (5).

The HABAM is a performance-based clinical bedside test, specially developed for use in older medical patients (7, 10). It includes the 3 sub-categories *balance, transfers and mobility*. The patient's mobility level is rated on a hierarchical scale of 22 items in the latest Rasch-refined HABAM version (8). It has sufficient face validity (5, 11) and can be administered within 2.6 ( $\pm 1$ ) min (11) without special equipment. However, in an older acute medical population, the HABAM has a ceiling effect, as approximately 25% of patients scored the maximum possible score at hospital admission (8, 12).

In 2008, the de Morton Mobility Index (DEMMI) was developed, based on the Rasch model, to overcome the limitations of existing measurement instruments of mobility in older people (13). The scale consists of 15 items that cover the broad mobility spectrum from bed-bound to high challenging balance abilities. The test can usually be administered in  $< 10$  min without additional equipment (13, 14). No floor or ceiling effects have been reported for its use in inpatient geriatric populations (13–15).

The HABAM and DEMMI are commonly used tools that may accurately measure mobility and facilitate the proposition of healthcare services in older medical patients. German translations and cross-cultural adaptations of both scales have been performed recently (14, 16). A comprehensive evaluation of the psychometric properties has been conducted for the German language DEMMI version in sub-acute geriatric rehabilita-

tion only (17). A comparative psychometric analysis could help clinicians and researchers to decide which instrument to use.

The aim of this study was to analyse the measurement properties of the DEMMI and the HABAM in an older acute medical inpatient population in a head-to-head comparison. In particular, this study aimed to determine the construct validity, test–retest reliability, measurement error and adequacy of scale width.

## METHODS

### Design and setting

This cross-sectional study examined the measurement properties of the DEMMI and HABAM in a consecutive sample of older acute medical patients in a geriatric hospital in Bochum, Germany. The study was approved by the Ethical Review board of the German Confederation of Physiotherapy (registration number: 2014–05) and *a priori* registered in the German Clinical Trials Register (DRKS00006231). All included participants provided written informed consent. The study used a follow-up measure to analyse responsiveness and aspects of interpretability, which will be reported in a future paper. In deviation from the registered protocol, inter-rater reliability was not assessed due to organizational constraints.

### Participants

All inpatients admitted to one of the acute geriatric wards of the hospital aged  $\geq 60$  years were eligible. Screening was performed over a period of 106 days during 2 recruitment phases between October 2014 and April 2015. Exclusion criteria were: blindness, deafness, severe dysphasia, language barriers, documented contra-indications for mobilization, physician-directed partial weight-bearing of the lower extremity, isolation for infection, impending death, coma or severely impaired vigilance, acute major organ failure or any acute psychiatric or medical/physical condition whereby mobility measurements could lead to a worsening of health state. Patients diagnosed with dementia (according to the medical chart) or an evident cognitive impairment (scored positive on Mini-Cog test (18)) were excluded.

### Procedures

**Validity sample.** Participants were assessed by a physiotherapist within the first 5 days after admission. The physiotherapist had basic work experience (young professional with 6 months of practical training in geriatric care) and was trained by clinical experts for a total of 6 h in use of the measurement instruments. In a single session, the DEMMI and HABAM were performed along with a comprehensive set of outcome measures. The test order was standardized, starting with the least physically challenging tests.

**Reliability sample.** To assess test–retest reliability, the DEMMI and HABAM were repeated by the same assessor in a sub-sample 1–6 h after the initial assessment in the same environment. Before the second test, participants were asked whether they had felt or experienced a change in their physical or mental condition (e.g. fatigue, pain, dizziness), and excluded if the answer was “yes”. For organizational reasons, recruitment for

the reliability sample started after approximately 50% ( $n=77$ ) of the target sample size was reached.

### Measurements

All assessments were performed in the patient’s hospital room and on the ward. Patients were scored at their highest level of safe function, using their usual walking aid. The same device was used for all assessments in a single session. Similar items in different assessments were only performed once, in order to reduce participant’s burden, e.g. standing with both feet together, which is required in the DEMMI and the Performance Oriented Mobility Assessment (POMA). For participants requiring some kind of physical assistance during ambulation (Functional Ambulation Categories (FAC) score  $\leq 2$ ), the walking tests were scored as “unable” (TUG, gait speed, 2-minute walk test) or “0 points” (POMA gait subscale), respectively.

The DEMMI is a bedside assessment, consisting of 15 hierarchical mobility items dealing with bed and chair mobility, ambulation, static and dynamic balance (13). The items are rated as 2- or 3-point response options. The ordinal raw score (0–19 points) can be transformed into the interval-level DEMMI score (0–100 points). Higher scores indicate better mobility. A German language version was used (14, 17).

The HABAM is a clinical bedside and interval-level mobility assessment that quantifies functional abilities in the sub-categories balance, transfers and ambulation (8). Higher scores indicate greater ability (0–26 points). A validated German HABAM translation was used (16).

The POMA is a well-established clinician-rated measure of mobility and fall risk (19). The maximum ordinal score of 28 points indicates better mobility.

The Rivermead Mobility Index (RMI) (20) is a cumulative index containing 15 mobility items, spanning a range of disabilities from being bedridden to running. In this study, the survey period was shortened and subjected to the prior 3 days as a meaningful worsening of general condition was expected in the days before hospital admission. Higher scores indicate better mobility.

The clinician-completed *Functional Ambulation Categories* distinguish 6 levels of walking ability subjected to the amount of assistance required over a walking distance of 10 m (21). Lower scores, where physical assistance is needed, indicate poor mobility.

The 2-minute walk test (2minWT) can be used to quantify functional-exercise capacity and walking ability (22). Subjects were asked to walk as far as possible within 2 min on the hospital corridor. Only 1 trial was performed, in order to avoid fatigue effects. For organizational reasons, no standardized walking track was used and the 2minWT was not performed until the first 50 patients had been assessed.

The Short Physical Performance Battery (SPPB) includes 3 objective tests of lower body function: a hierarchical test of standing balance, a 4-m walk test (4mWT) and 5 times chair rise test ( $5 \times CRT$ ) (23). The SPPB has a scoring from 0 (unable) to 4 points for every sub-test, with a maximum of 12 points in total (ordinal scaled).

Habitual gait speed (in m/s) over a distance of 4 m was assessed as part of the SPPB. Timing was started when the participant began walking. The shorter time of 2 trials was used for analysis (24).

For the  $5 \times CRT$ , participants were asked to stand up and sit down from a chair (standardized height: 46 cm) as fast as possible 5 times, with their hands crossed in front of their chest.

The TUG is a performance-based test that assesses basic mobility functions. The patient is asked to stand up from a chair

(height 46 cm), walk 3 m, turn around, walk back and return to the chair (25). In the present study, a cone had to be encircled. Participants chose the turning side. A familiarization trial was followed by one counted trial.

To assess *self-rated mobility*, a short ICF definition of mobility was given. Patients then rated their current level of mobility on a 5-point rating scale (very bad, bad, normal/moderate, good, very good; -2 to +2). Self-rated mobility scores were used to describe the sample only due to the uncertain validity of these scores.

The *Mini Mental State Examination* (MMSE 0–30 points) (26) is a global measure of cognitive functioning.

The *Barthel Index* (27) (BI; 0–100 points) is a measure of independence in the activities of daily living (ADL).

The MMSE and BI were recorded by the nursing and occupational therapy staff as part of routine care at another time than the DEMMI/HABAM. Thus, MMSE and BI were only used to describe the sample. In both scales, higher scores indicate better functioning.

#### Statistical analysis

Data were analysed using SPSS statistical software (SPSS 21.0 for Windows; IBM Corp.; Armonk, NY, USA). Interval-based data were examined for normal distribution with the Shapiro-Wilk test of normality and by visual inspection of the related histograms and P-P-plots.  $p < 0.05$  indicated statistical significance.

#### Measurement properties

**Construct validity.** In the absence of a gold standard for mobility, construct validity was assessed by following the methodological approach of hypotheses testing (28, 29). A target sample size of at least 100 was set according to the “COnsensus based Standards for the selection of health Measurement INstruments” (COSMIN) (30).

Aspects of convergent and known-groups validity (other functional assessments and clinical information) were used to formulate 13 hypotheses (H1 to H13) (31). All hypotheses were formulated *a priori*, based on existing literature and expertise of clinicians and the research team. One-tailed Spearman’s rho analysis was applied because directions of the correlations were hypothesized *a priori* and all data were either ordinal or not normally distributed. For tests in which lower scores represent better functioning (TUG and 5 × CRT), a negative correlation was hypothesized. Expected strengths of correlations are reported unidirectionally to improve readability.

For each hypothesis on known-groups validity, we expected a statistically significant difference between the mean scores of both groups, respectively. A 1-sided Mann–Whitney *U* test for independent samples was used to compare groups as directional hypotheses were formulated *a priori*.

- H1-6: DEMMI/HABAM scores correlate strongly ( $>0.7$ ) with other broad measures of mobility (DEMMI/HABAM, RMI, POMA, FAC, SPPB, TUG) as reported by others (13, 17, 32).
- H7: DEMMI/HABAM scores correlate strongly ( $>0.7$ ) with the 2minWT, a measure of walking capacity. De Morton et al. (33) found  $\rho = 0.76$  between DEMMI and the 6-min walk test in older patients with hip fracture.
- H8: DEMMI/HABAM scores correlate moderately ( $0.5 < \rho \leq 0.7$ ) with gait speed, a single component mobility measure. De Morton et al. (33) found  $\rho = 0.62$  between DEMMI scores and gait speed of older patients with hip fracture.
- H9: DEMMI/HABAM scores correlate moderately ( $0.5 < \rho \leq 0.7$ ) with the 5 × CRT, a single component mobility

measures. Jans et al. (32) found  $\rho = -0.63$  between DEMMI and 5 × CRT in older patients with knee or hip osteoarthritis.

- H10: Ambulatory participants ( $FAC \geq 3$ ) ambulating without a walking aid have higher DEMMI/HABAM scores than participants using a walking aid.
- H11: Independent walkers ( $FAC \geq 4$ ) have higher DEMMI/HABAM scores than non-ambulatory participants or dependent walkers ( $FAC \leq 3$ ).
- H12: Participants who are able to perform the TUG (capturing basic mobility aspects) have higher DEMMI/HABAM scores than participants who cannot perform the TUG.
- H13: Item 14 of the RMI assesses the ability to climb 4 steps of a stairway, up and down, without a handrail, but with the use of a walking aid. We expected participants who scored positively and those who scored negatively on this RMI item to show a difference in mean DEMMI/HABAM scores.

We decided against defining an *a priori* threshold of a percentage of hypotheses (e.g. 75% as applied by Dobson et al. (34)) that would need to be confirmed in order for a measurement instrument to be considered valid (29). Along with others (35), we do not consider that the broad concept of construct validity can be judged “good” or “bad” according to an arbitrary threshold of confirmed hypotheses of varying importance. Instead, we leave it to the reader to decide the percentage of confirmed hypotheses she or he deems acceptable.

#### Reliability

**Internal consistency reliability.** Cronbach’s alpha as an adequate measure of internal consistency in unidimensional scales (29) was derived solely from the baseline data of the validity sample for the DEMMI because of the larger sample size. A Cronbach’s alpha between 0.7 and 0.95 was considered acceptable (29). For the HABAM, an internal consistency reliability analysis could not be performed, as no item-level data was available due to the hierarchical structure.

**Test–retest reliability.** Test–retest reliability was examined using the intra-class correlation coefficient (ICC) model 2,1 (2-way random effects model; ICC<sub>AGREEMENT</sub>) (31, 36). Variance components were estimated using the VARCOMP tool in SPSS. An ICC of  $\geq 0.7$  or higher was deemed acceptable (29). According to COSMIN, we required at least 30 participants (30).

**Agreement: Measurement error.** The standard error of measurement (SEM<sub>AGREEMENT</sub>) was calculated using the same variance components used for the ICC<sub>AGREEMENT</sub> calculation (31).

**Agreement: Bland and Altman plot.** Bland and Altman plots with corresponding bar charts were used to illustrate agreement for the test–retest condition (37). The 95% limits of agreement require homoscedasticity and normally distributed differences (38). A positive Kendall’s tau ( $\tau$ ) correlation between the absolute differences and the corresponding means (39)  $> 0.1$  was deemed to denote heteroscedasticity. In case of heteroscedastic data, the following formula was used to calculate the limits of agreement:

$$-2X \frac{(10^a - 1)}{(10^a + 1)} \text{ and } +2X \frac{(10^a - 1)}{(10^a + 1)},$$
 with  $a = 95\%$  limits of agreement of the 10 log-transformed data and  $X$  the mean score (40).

#### Interpretability

**Minimal detectable change.** The minimal detectable change (MDC) with 90% and 95% confidence was calculated based

on the test–retest reliability data as  $MDC_{90} = 1.64 * \sqrt{2} * SEM_{AGREEMENT}$  and  $MDC_{95} = 1.96 * \sqrt{2} * SEM_{AGREEMENT}$  respectively.

**Floor and ceiling effects.** A floor or ceiling effect was considered if  $\geq 15\%$  of the study participants scored the highest or lowest possible DEMMI/HABAM score or within the MDC of the extremes, respectively (29).

## RESULTS

A total of 158 acute medical patients were included in this study, 30 of whom were re-assessed for test–retest analysis (participant flow: Fig. 1). Validity sample participants had a mean age of 81 years, 68% were female, and 51% presented with a musculoskeletal disorder (validity and reliability sample characteristics: Table I). Table II shows the mobility-related outcomes of the validity sample at hospital admission.

The 2minWT was not applied in 50 (32%) participants, and there were 32 (30%) of 108 assessed participants who were not physically able to perform the 2minWT. There were further floor effects in some other tests as some participants could not perform the gait speed measure ( $n=51$ ; 32%), the  $5 \times CRT$  ( $n=119$ ; 75%) and the TUG ( $n=58$ ; 37%) due to insufficient balance, walking and sit-to-stand transfer abilities.

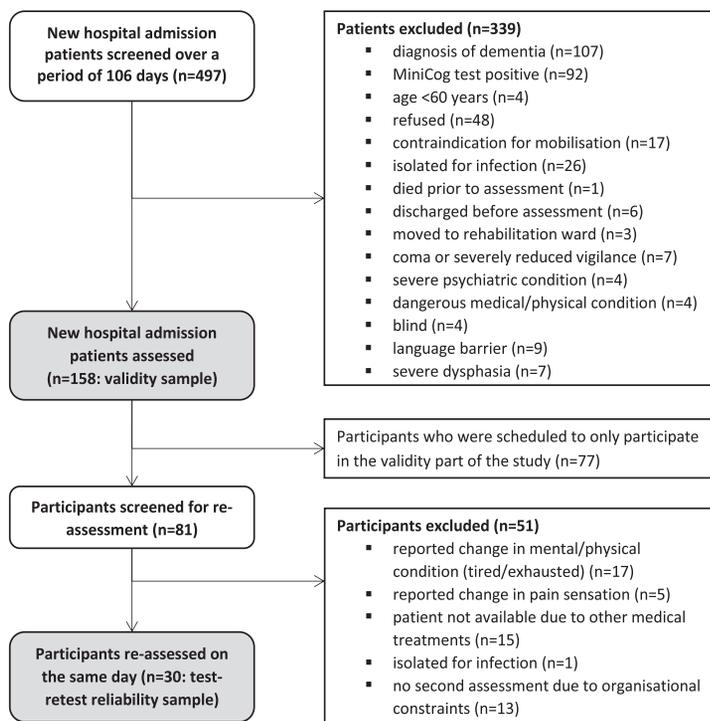
### Distribution of scores

DEMMI baseline scores ( $n=158$ ) ranged from 0 to 85 out of 100 and were normally distributed (Shapiro-

**Table I.** Characteristics of the overlapping validity and reliability samples of participants (total  $n=158$ ) at admission

|   | Validity sample<br>( $n=158$ ) | Test-retest<br>sample ( $n=30$ ) |
|---|--------------------------------|----------------------------------|
| Age, years, mean (SD) [range]   | 81 (7) [63–95]                 | 81 (8) [65–94]                   |
| Women, $n$ (%)  | 107 (68)                       | 17 (57)                          |
| Primary diagnosis according to ICD-10 categories, $n$ (%)             |                                |                                  |
| Musculoskeletal   | 81 (51)                        | 12 (40)                          |
| Circulatory   | 16 (10)                        | 2 (7)                            |
| Nervous system  | 15 (10)                        | 4 (13)                           |
| Respiratory   | 15 (10)                        | 3 (10)                           |
| Other   | 31 (20)                        | 9 (30)                           |
| Distribution of primary disease classification at assessment, $n$ (%) |                                |                                  |
| Chronic condition needing acute medical care (> 3 months)             | 17 (11)                        | 4 (13)                           |
| Acute condition (< 3 months)  | 116 (73)                       | 21 (70)                          |
| Missing/unclear   | 25 (16)                        | 5 (17)                           |
| Duration of acute condition at assessment, days, mean (SD) [range]    | 21 (15) [3–86]                 | 16 (7) [8–36]                    |
| Time between admission and assessment, days, mean (SD) [range]        | 3 (1) [1–6]                    | 3 (1) [1–5]                      |
| In-hospital walking aid (self-reported), $n$ (%)                      |                                |                                  |
| None  | 13 (8)                         | 5 (17)                           |
| Cane/single crutch  | 9 (6)                          | 0 (0)                            |
| Rollator  | 94 (60)                        | 19 (63)                          |
| Other   | 10 (6)                         | 2 (7)                            |
| Wheelchair (non-ambulatory)   | 32 (20)                        | 4 (13)                           |
| Ambulation, $n$ (%)   |                                |                                  |
| Independent walkers (FAC $\geq 4$ )                                   | 68 (43)                        | 20 (67)                          |
| Stair-climbing abilities, $n$ (%)                                     |                                |                                  |
| Able to climb stairs  | 34 (22)                        | 14 (47)                          |
| Self-rated mobility, points, median (IQR)                             | 0 (–1–0)                       | 0 (–1–0)                         |
| Mini Mental State Examination   |                                |                                  |
| Valid, $n$ (%)  | 139 (88)                       | 28 (93)                          |
| Score, points, median (IQR)   | 27 (26–29)                     | 28 (27–29)                       |
| Barthel Index   |                                |                                  |
| Valid, $n$ (%)  | 157 (99)                       | 30 (100)                         |
| Score, points, median (IQR)   | 40 (35–50)                     | 45 (35–50)                       |

SD: standard deviation; ICD-10: International Classification of Diseases 10<sup>th</sup> version; FAC: Functional Ambulation Categories. IQR: interquartile range.



**Fig. 1.** Flow chart of study participants.

Wilk:  $W=0.99$ ;  $p=0.08$ ) (Fig. 2a). HABAM baseline scores ranged from 0 to 26 points out of 26 and were not normally distributed (Shapiro-Wilk:  $W=0.88$ ;  $p<0.001$ ) (Fig. 2b).

### Construct validity

DEMMI and HABAM scores correlated with other mobility related outcomes, and differed between known clinical groups (Table II). For each instrument, 11 out of 13 hypotheses (85%) could be confirmed.

### Reliability

**Internal consistency reliability.** The Cronbach's alpha of the DEMMI was 0.87, which implies good internal consistency. For the HABAM, internal consistency reliability analysis could not be performed.

**Test–retest reliability.** Thirty participants (characteristics shown in Table I) were asses-

**Table II.** Construct validity of the de Morton Mobility Index (DEMMI) and the Hierarchical Assessment of Balance And Mobility (HABAM) ( $n = 158$ )

| No. Hypothesis  | Comparison measurement instrument        | DEMMI  |  | HABAM  |     | Hypothesis confirmed | Rho                    | 95% CI         | Hypothesis confirmed |
|---|--|--|--|--|-----|----------------------|------------------------|----------------|----------------------|
|   |  | Mean (SD) or median (IQR)                                | Observed correlation with DEMMI (Spearman's correlation) | Observed correlation with HABAM (Spearman's correlation) |     |                      |                        |                |                      |
| 1 Concerning 1-7, a correlation of $>0.7$ was expected between the DEMMI/HABAM, and other broad measures of mobility scales and walking endurance                       | DEMMI                                    | 44.1 (15.9)  | na   | na   | na  | Yes                  | 0.87                   | 0.83 to 0.90   | Yes                  |
|   | HABAM                                    | 17.2 (5.4)   | 0.87   | 0.83 to 0.90   | na  | Yes                  | na                     | na             | na                   |
| 2 Rivermead Mobility Index  | Rivermead Mobility Index                 | 7 (5-9)  | 0.92   | 0.89 to 0.94   | Yes | Yes                  | 0.85                   | 0.80 to 0.89   | Yes                  |
|   | Performance Oriented Mobility Assessment | 15 (6-20)  | 0.92   | 0.89 to 0.94   | Yes | Yes                  | 0.86                   | 0.81 to 0.90   | Yes                  |
| 3 Functional Ambulation Categories  | Functional Ambulation Categories         | 3 (2-4)  | 0.87   | 0.83 to 0.90   | Yes | Yes                  | 0.86                   | 0.81 to 0.90   | Yes                  |
|   | Short Physical Performance Battery       | 3 (1-6)  | 0.91   | 0.88 to 0.93   | Yes | Yes                  | 0.86                   | 0.81 to 0.90   | Yes                  |
| 4 Timed Up And Go test ( $n = 100$ )  | Timed Up And Go test ( $n = 100$ )       | 23 (10)  | -0.67  | -0.76 to -0.55   | No  | No                   | -0.60                  | -0.71 to -0.46 | No                   |
|   | 2 min walk test ( $n = 76$ )             | 57 (27)  | 0.76   | 0.65 to 0.84   | Yes | Yes                  | 0.75                   | 0.63 to 0.83   | Yes                  |
| 5 A moderate correlation ( $0.5 < \rho \leq 0.7$ ) was expected between the DEMMI/HABAM and other single-component mobility scales                                      | Gait speed ( $n = 107$ )                 | 0.59 (0.20)  | 0.62   | 0.49 to 0.72   | Yes | Yes                  | 0.61                   | 0.48 to 0.72   | Yes                  |
|   | 5 x chair rise test ( $n = 39$ )         | 22 (10)  | -0.46  | -0.67 to -0.18   | No  | No                   | -0.38                  | -0.61 to -0.09 | No                   |
| Observed mean DEMMI/HABAM scores (points) according to clinical groups  |  | Statistical significance (Mann-Whitney $U$ test, 1-fold) |  | Statistical significance (Mann-Whitney $U$ test, 1-fold) |     |                      |                        |                |                      |
| Clinical groups   |  | DEMMI  | HABAM  |  |     |                      |                        |                |                      |
| 10 A statistically significant mean difference between ambulatory ( $FAC \geq 3$ ; $n = 108$ ) participants walking without vs participants walking with a walking aid. | No walking aid ( $n = 13$ )              | 65.9 (12.2)  | 24.4 (2.9)   | $U = 191$ ; $p < 0.01$                                   |     | Yes                  | $U = 138$ ; $p < 0.01$ |                | Yes                  |
|   | Walking aid ( $n = 95$ )                 | 50.0 (9.4)   | 19.1 (2.6)   |  |     |                      |                        |                |                      |
| 11 A statistically significant mean difference between independently ambulatory ( $FAC \geq 4$ ) vs dependently ambulatory/non-ambulatory ( $FAC < 4$ ) participants.   | Independent walkers ( $n = 68$ )         | 57.2 (9.3)   | 21.5 (2.0)   | $U = 321$ ; $p < 0.01$                                   |     | Yes                  | $U = 115$ ; $p < 0.01$ |                | Yes                  |
|   | Dependent/non-ambulatory ( $n = 90$ )    | 34.1 (12.3)  | 13.9 (4.8)   |  |     |                      |                        |                |                      |
| 12 A statistically significant mean difference between participants who can perform the TUG and those who are not able to perform the TUG.                              | TUG possible ( $n = 100$ )               | 53.2 (10.4)  | 20.1 (2.8)   | $U = 111$ ; $p < 0.01$                                   |     | Yes                  | $U = 508$ ; $p < 0.01$ |                | Yes                  |
|   | TUG not possible ( $n = 58$ )            | 28.3 (10.7)  | 12.2 (5.0)   |  |     |                      |                        |                |                      |
| 13 A statistically significant mean difference between participants who can climb stairs and those who cannot (RMI Item 14).  | Able to climb stairs ( $n = 34$ )        | 59.7 (10.3)  | 22.1 (2.5)   | $U = 542$ ; $p < 0.01$                                   |     | Yes                  | $U = 517$ ; $p < 0.01$ |                | Yes                  |
|   | Not able to climb stairs ( $n = 124$ )   | 39.8 (14.5)  | 15.9 (5.2)   |  |     |                      |                        |                |                      |

DEMMI: de Morton Mobility Index; HABAM: Hierarchical Assessment of Balance And Mobility; TUG: Timed Up And Go test; FAC: Functional Ambulation Categories; RMI: Rivermead Mobility Index; SD: standard deviation; IQR: interquartile-range; CI: confidence interval; na: not applicable.

sed on a second occasion by the same assessor 3.5 (SD 1.5) (range 1-6) h after baseline assessment.

For the DEMMI ( $p = 0.96$ ) and the HABAM ( $p = 0.33$ ) there were neither statistically significant mean test-retest differences, nor considerable variance due to systematic differences over time (Table III). The ICC<sub>AGREEMENT</sub> of the DEMMI and HABAM of 0.98 and 0.99 indicated excellent test-retest reliability.

**Agreement: measurement error.** The SEM<sub>AGREEMENT</sub> was 2.34 points for the DEMMI and 0.39 points for the HABAM.

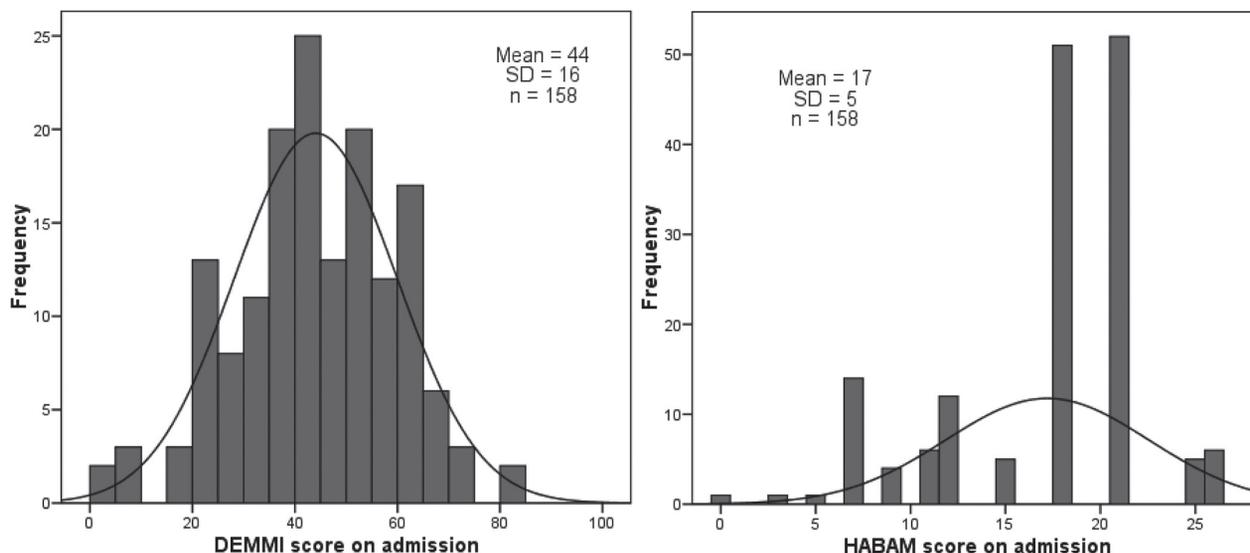
**Agreement: Bland and Altman plot.** DEMMI data showed heteroscedasticity ( $\tau = 0.26$ ). This finding

was reflected by 95% limits of agreement becoming wider, meaning a higher variability between baseline and retest with increasing values. For any X, the limits of agreement were  $-0.09X - 0.03$  to  $+0.09X - 0.03$  (Fig. 3a).

Data of 2 HABAM measures was homoscedastic ( $\tau = -0.09$ ), so that constant limits of agreement were  $-1.0$  to  $1.2$  (Fig. 3b).

### Interpretability

**Minimal detectable change.** MDC values of the DEMMI were  $MDC_{90} = 5.4$  and  $MDC_{95} = 6.5$ . As the 100-point DEMMI scale consists of whole numbers



**Fig. 2.** (a) De Morton Mobility Index (DEMMI) scores on admission to the acute geriatric hospital. Two participants (1.2%) scored 0 points and 0 participants (0%) scored 100 points. (b) Hierarchical Assessment of Balance and Mobility (HABAM) scores on admission to the acute geriatric hospital. One participant (0.6%) scored 0 points and 6 participants (3.8%) scored 26 points.

only, a DEMMI change must be at least 6 points to have 90% confidence that this change is beyond measurement error and at least 7 points to have 95% confidence.

MDC values of the HABAM were  $MDC_{90} = 0.9$  and  $MDC_{95} = 1.1$ . As the 26-point HABAM scale consists of whole numbers only, a change score of 1 point

gives 90% confidence that this is a “real” change, and a change score of 2 points produces 95% confidence.

*Floor and ceiling effects.* Figs 2a and 2b, together with Table IV, illustrate that, subjected to the  $MDC_{95}$ -ranges, for neither of both scales any absolute floor or ceiling effects occurred.

**Table III.** Test-retest reliability of the de Morton Mobility Index (DEMMI) and the Hierarchical Assessment of Balance and Mobility (HABAM) ( $n = 30$ )

|   | DEMMI              | HABAM             |
|---|--------------------|-------------------|
| Score occasion 1, mean (SD)                                 | 52 (16)            | 20 (5)            |
| Score occasion 2, mean (SD)                                 | 52 (16)            | 20 (5)            |
| Mean difference (95% CI)                                    | -0.03 (-1.27-1.20) | 0.10 (-0.11-0.31) |
| Variance between patients                                   | 265.207            | 24.161            |
| Variance due to systematic differences between observations | < 0.001            | < 0.001           |
| Residual variance   | 5.465              | 0.150             |
| Intraclass correlation coefficient ( $ICC_{AGREEMENT}$ )    | 0.98 (0.96-0.99)   | 0.99 (0.99-0.99)  |
| Standard error of measurement ( $SEM_{AGREEMENT}$ )         | 2.34               | 0.39              |

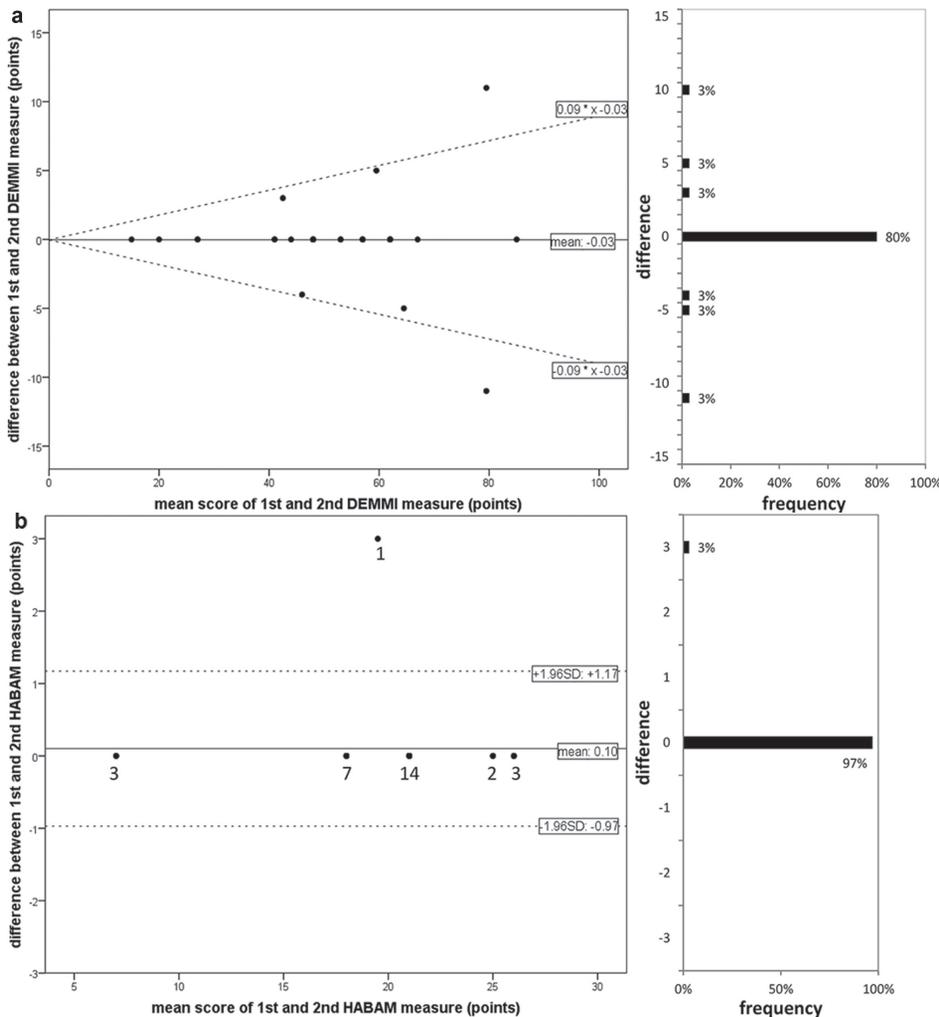
SD: standard deviation; CI: confidence interval.

**Table IV.** Floor and ceiling effects of the de Morton Mobility Index (DEMMI) and the Hierarchical Assessment of Balance and Mobility (HABAM)

| Scale width                | Absolute floor | Absolute ceiling | $MDC_{95}$ | $MDC_{95}$ -range from |         | % of patients scoring within $MDC_{95}$ -range |         |
|----------------------------|----------------|------------------|------------|------------------------|---------|--|---------|
|                            |                |                  |            | Floor                  | Ceiling | Floor  | Ceiling |
| DEMMI (0-100) <sup>a</sup> | No (1%)        | No (0%)          | 6.5        | 0-7                    | 83-100  | No (1%)  | No (1%) |
| HABAM (0-26) <sup>a</sup>  | No (1%)        | No (4%)          | 1.1        | 0-2                    | 24-26   | No (1%)  | No (7%) |

From left to right, for both scales, the absolute scale width and the absolute floor and ceiling effects are presented. Followed by the minimal detectable change ( $MDC_{95}$ ) and the percentage of scores that fell within the  $MDC_{95}$ -range for both extremes.

<sup>a</sup>Ranges reflect lower to high mobility functions.



**Fig. 3.** (a) Bland-Altman plot of the test-retest reliability of the de Morton Mobility Index (DEMMI). (b) Bland-Altman plot of the test-retest reliability of the Hierarchical Assessment of Balance and Mobility (HABAM) (numbers indicate frequencies of participants). In each figure, the x-axis represents the mean scores of 2 observations made by the same assessor and the y-axis represents the difference between the 2 observations. The straight line represents the mean difference between both measures; dotted lines represent the 95% upper and lower limits of agreement. The bar chart on the right side illustrates the frequency of differences between the 2 measures.

**DISCUSSION**

This is the first study to examine the reliability and validity of the German language versions of the DEMMI and HABAM in an acute geriatric hospital setting. Both scales seem to be sufficiently and equally construct-valid and test-retest reliable. These results support the use of both the DEMMI and HABAM as measures of mobility in older acute medical patients.

Our sample seems quite comparable to the patients analysed in the original DEMMI and HABAM studies (9, 12, 13) with respect to the HABAM score, age, and gender distribution, even though the DEMMI score was somehow lower than in the original validation study (44 (SD 16) vs 52 (SD 21) points) (12, 13). This deviation might be explained by the different sample compositions (51% musculoskeletal, mostly fall-related acute conditions in the current study vs 38% cardio-respiratory principal diagnoses in the original validation study) (13).

The point correlation between the DEMMI and HABAM, which are considered to be the most valid

instruments of older people’s mobility (5, 15), was 0.87. This can be interpreted as a high correlation, indicating good construct-validity of both scales.

The lower than expected point correlations between DEMMI and HABAM scores and TUG scores ( $\rho = -0.67$  and  $\rho = -0.60$ , respectively) as well as 5 × CRT times ( $\rho = -0.46$  and  $\rho = -0.38$ ) might be attributed to the lower sample size (TUG:  $n = 100$  and 5 × CRT:  $n = 39$ ). Studies comparing the HABAM to the TUG or 5 × CRT are lacking. Other authors found correlations between DEMMI and TUG of  $-0.48$ ,  $-0.57$  and  $-0.73$  in sub-acute geriatric patients (41), patients with Parkinson’s disease (42) and older patients with knee or hip osteoarthritis (32), respectively. Jans et al. (32) reported a point correlation between the DEMMI and chair rise time of  $-0.63$ . Since the 95% CIs of the correlations found in our study include the expected point correlations, these falsified hypotheses do not raise major concerns. Besides, 85% of our hypotheses were confirmed, and according to the COSMIN recommendations (29), this indicates sufficient construct-validity.

Test–retest reliability ( $ICC_{AGREEMENT}$  of 0.98 and 0.99) and internal consistency reliability (DEMMI only, Cronbach's  $\alpha=0.87$ ) were excellent and comply with quality criteria proposed by COSMIN (29). Others have found comparable, but slightly lower, intra-rater reliability (0.86–0.92) (43). It is likely that participants were less stable in the de Morton et al. (43) study than in our study, since patients who fared “bit worse” or a “bit better” at discharge were still considered unchanged. Rockwood et al. (9) reported an ICC of 0.91 (95% CI 0.85–0.94) for the Rasch-refined HABAM version in 63 geriatric inpatients assessed within 2 hospital days by a geriatrician. The DEMMI and HABAM appear to produce stable results over a short period of time. However, this requires physiotherapists who are highly familiar with these instruments.

The  $MDC_{90}$  values of 6 points (DEMMI) and 1 point (HABAM) found in the present study represent 6% of the DEMMI and approximately 4% of the HABAM scale width, respectively. This measurement error appears to be relatively small. Other authors have reported DEMMI  $MDC_{90}$  scores of 8–10 points (13, 17, 41, 43). An  $MDC$  of the Rasch refined HABAM has not been published. Again, the small measurement error might be to the inclusion and re-assessment only of stable patients over a very short period of time. Furthermore, the rater was extensively trained in the conduction of the measurement instruments and not blinded towards the results of the first assessment. However, we assume that the DEMMI and HABAM can produce reliable scores even by unexperienced raters if similarly comprehensive training procedures are followed.

As expected, and reported by others previously (5, 33), measures that require the patient to stand, walk and transfer from sit to stand, showed significant floor effects. In 30% (2minWT) to 75% ( $5 \times CRT$ ), patients were not able to perform these tests at hospital admission. Compared with the TUG, gait speed,  $5 \times CRT$  and 2minWT, which have limitations, the DEMMI and HABAM have a more inclusive scale width and showed neither floor nor ceiling effects. However, only the DEMMI scores showed a normal distribution, due to the broader scale width.

### Study strengths and limitations

A strength of this study is the sufficiently large (30) validity sample size of 158 participants. The sample size of 30 participants for the test–retest part of the study can be judged as “fair” (28, 29), and is in line with previous examinations (9, 13, 32). The lower limit of the 95% CI of the ICC of the DEMMI and HABAM was higher than the recommended minimum standard of reliability ( $ICC \geq 0.70$ ) (29).

Despite the large sample sizes and consecutive recruitment of multi-morbid participants with a wide spectrum of diseases, the external validity of this study might be limited, as data was collected in only one hospital. While participants might have become slightly fatigued during the assessment session, the test order was kept constant throughout the study so that this should have a limited impact on the current results. Furthermore, 199 participants with dementia ( $n=107$ ) or other cognitive impairment ( $n=92$ ) were excluded, which amounts to 40% of all inpatients admitted to the hospital during the recruitment period. The clinimetric properties of the DEMMI and HABAM in older acute medical patients with cognitive impairment remain unclear and should be evaluated in further studies.

In the absence of a gold standard, a comprehensive set of mobility reference assessments was used to assess construct validity based on multiple *a priori* hypotheses. This allowed us to take into account the broad spectrum of mobility covered by the DEMMI and HABAM. We considered these instruments sufficiently valid. RMI and FAC are usually applied in neurorehabilitation (20, 21), but there are some arguments for the (construct) validity of these scales as measures of mobility in older acute medical patients. We included multi-component measures of walking and chair transfers, such as gait speed, TUG and 2minWT, because these instruments are frequently used in (German) geriatric rehabilitation. The expected floor effects, which confirm the limitations of these instruments (5, 14, 44), might have biased the point correlations used for the construct-validity analysis.

The external validity of the test–retest reliability estimations might be limited to physiotherapists with a comparably basic work experience who are not blinded towards the results of the first assessment.

Following Stratford et al. (45), the calculation of the  $MDC$  should rely on data of stable patients (longitudinal approach) (31, 45). In clinical practice, it is unusual to have 1–6 h between assessments only, as patients are typically assessed every couple of days, or at hospital admission and discharge only. However, a more frequent mobility assessment has been proposed (3) to track illness and recovery of older patients in the acute hospital. Daily bedside observation of mobility by means of the HABAM allow for a more thorough assessment of acute changes in the health of older people (3). Further research should also generate reliability data over a longer period of time, such as 18 days, which was the study participant's mean length of stay. It may be challenging to find stable patients, however, because changes in health and function become more likely with increasing periods of time.

The responsiveness to change, the minimal important change and a Rasch analysis have not been

addressed in the current study. These analyses will be performed, as described in the registered study protocol, but will be reported in another paper. However, other studies (7, 8, 13) showed sufficient evidence for the DEMMI and the HABAM to be responsive and unidimensional measures of mobility in older acute medical patients.

### Clinical implications

This study generates evidence that the German language versions of the DEMMI and HABAM have adequate measurement properties in acute older medical inpatients. Validity, reliability and MDC values are in good agreement with other studies. Thus, the habit of using single-component instruments, such as ambulation or lower limb strength, in acute geriatric care, which have large floor effects and measure only single aspects of mobility, should further be cautiously reconsidered. The DEMMI and HABAM operate without special equipment and produce results within 3–10 min. While the HABAM is less time-consuming, the DEMMI provides a more readily understandable score range (0–100 points). To the best of our knowledge, in this or a previous study, no minor or major adverse events have occurred during the performance of any assessment of DEMMI or HABAM, thus both instruments seem safe to administer (7–11, 13–15). By simple observation of physical performance, both instruments identify changes in mobility over the whole clinical spectrum, which are relevant to the patient, therapist, and healthcare provider.

### Conclusion

The DEMMI and HABAM appear to be almost equally suitable to measure mobility in older acute medical patients. Since only the DEMMI scores were normally distributed in this study, and ceiling effects have been reported for the HABAM in other studies (8, 12), the DEMMI seems to have a slight advantage over the HABAM.

## ACKNOWLEDGEMENTS

The authors thank all participants for participating in this study, and further acknowledge the support of the physiotherapy, occupational therapy, nursing and medical staff of the Marien-Hospital Wattenscheid, Bochum.

*The authors have no conflicts of interest to declare.*

## REFERENCES

- Ostir GV, Berges IM, Ottenbacher KJ, Fisher SR, Barr E, Hebel JR, et al. Gait speed and disability in older adults. *Arch Phys Med Rehabil* 2015; 96: 1641–1645.
- Graham JE, Fisher SR, Berges I-M, Kuo Y-F, Ostir GV.

- Walking speed threshold for classifying walking independence in hospitalized older adults. *Phys Ther* 2010; 90: 1591–1597.
- Hubbard RE, Eeles EMP, Rockwood MRH, Fallah N, Ross E, Mitnitski A, et al. Assessing balance and mobility to track illness and recovery in older inpatients. *J Gen Intern Med* 2011; 26: 1471–1478.
- World Health Organization. International classification of functioning, disability and health: ICF. Geneva: World Health Organization; 2001.
- de Morton NA, Berlowitz DJ, Keating JL. A systematic review of mobility instruments and their measurement properties for older acute medical patients. *Health Qual Life Outcomes* 2008; 6: 44.
- Davenport SJ, Paynter S, de Morton NA. What instruments have been used to assess the mobility of community-dwelling older adults? *Phys Ther Rev* 2008; 13: 345–354.
- MacKnight C, Rockwood K. A hierarchical assessment of balance and mobility. *Age Ageing* 1995; 24: 126–130.
- MacKnight C, Rockwood K. Rasch analysis of the hierarchical assessment of balance and mobility (HABAM). *J Clin Epidemiol* 2000; 53: 1242–1247.
- Rockwood K, Rockwood MRH, Andrew MK, Mitnitski A. Reliability of the hierarchical assessment of balance and mobility in frail older adults. *J Am Geriatr Soc* 2008; 56: 1213–1217.
- MacKnight C, Rockwood K. Mobility and balance in the elderly: a guide to bedside assessment. *Postgrad Med* 1996; 99: 269–271, 275–276.
- MacKnight C, Sibley A, Rockwood K. The sensibility of bedside tests of balance and mobility. *Geriatrics Today* 2002; 5: 140–143.
- de Morton NA, Davidson M, Keating J. Validity, responsiveness and the minimal clinically important difference for the de Morton Mobility Index (DEMMI) in an older acute medical population. *BMC Geriatr* 2010; 10: 72.
- de Morton NA, Davidson M, Keating JL. The de Morton Mobility Index (DEMMI): an essential health index for an ageing world. *Health Qual Life Outcomes* 2008; 6: 63.
- Braun T, Schulz R-J, Hoffmann M, Reinke J, Tofaute L, Urner C, et al. German version of the de Morton Mobility Index. First clinical results from the process of the cross-cultural adaptation. *Z Gerontol Geriatr* 2015; 48: 154–163.
- de Morton NA, Nolan J, O'Brien M, Thomas S, Govier A, Sherwell K, et al. A head-to-head comparison of the de Morton Mobility Index (DEMMI) and Elderly Mobility Scale (EMS) in an older acute medical population. *Disabil Rehabil* 2015; 37: 1881–1887.
- Braun T, Rieckmann A, Grüneberg C, Marks D, Thiel C. Hierarchical assessment of balance and mobility. *Z Gerontol Geriatr* 2016; 49: 386–397.
- Braun T, Schulz R-J, Reinke J, van Meeteren NL, de Morton NA, Davidson M, et al. Reliability and validity of the German translation of the de Morton Mobility Index (DEMMI) performed by physiotherapists in patients admitted to a sub-acute inpatient geriatric rehabilitation hospital. *BMC Geriatr* 2015; 15: 1660.
- Milian M, Leiherr A-M, Straten G, Müller S, Leyhe T, Eschweiler GW. The Mini-Cog versus the Mini-Mental State Examination and the Clock Drawing Test in daily clinical practice: screening value in a German Memory Clinic. *Int Psychogeriatr* 2012; 24: 766–774.
- Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc* 1986; 34: 119–126.
- Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. *Int Disabil Stud* 1991; 13: 50–54.
- Holden MK, Gill KM, Magliozzi MR, Nathan J, Piehl-Baker L. Clinical gait assessment in the neurologically impaired. Reliability and meaningfulness. *Phys Ther* 1984; 64: 35–40.
- Pin TW. Psychometric properties of 2-minute walk test:

- a systematic review. *Arch Phys Med Rehabil* 2014; 95: 1759–1775.
23. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 1994; 49: 85–94.
  24. Graham JE, Ostir GV, Fisher SR, Ottenbacher KJ. Assessing walking speed in clinical research: a systematic review. *J Eval Clin Pract* 2008; 14: 552–562.
  25. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991; 39: 142–148.
  26. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189–198.
  27. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965; 14: 61–65.
  28. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63: 737–745.
  29. Terwee CB, Bot SDM, Boer MR de, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34–42.
  30. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, Vet HCW de. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21: 651–657.
  31. Vet HCW de, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*. Cambridge, New York: Cambridge University Press; 2011.
  32. Jans MP, Slootweg VC, Boot CR, de Morton NA, van der Sluis G, van Meeteren NL. Reproducibility and validity of the Dutch translation of the de Morton Mobility Index (DEMMI) used by physiotherapists in older patients with knee or hip osteoarthritis. *Arch Phys Med Rehabil* 2011; 92: 1892–1899.
  33. de Morton NA, Harding KE, Taylor NF, Harrison G. Validity of the de Morton Mobility Index (DEMMI) for measuring the mobility of patients with hip fracture during rehabilitation. *Disabil Rehabil* 2013; 35: 105–111.
  34. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis Cartil* 2012; 20: 1548–1562.
  35. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013; 22: 1889–1905.
  36. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1: 30–46.
  37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
  38. Altman DG, Bland JM. Measurement in Medicine: the analysis of method comparison studies. *The Statistician* 1983; 32: 307–317.
  39. Brehm MA, Scholtes VA, Dallmeijer AJ, Twisk JW, Harlaar J. The importance of addressing heteroscedasticity in the reliability analysis of ratio-scaled variables: an example based on walking energy-cost measurements. *Dev Med Child Neurol* 2012; 54: 267–273.
  40. Euser AM, Dekker FW, Le Cessie S. A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. *J Clin Epidemiol* 2008; 61: 978–982.
  41. de Morton NA, Lane K. Validity and reliability of the de Morton Mobility Index in the subacute hospital setting in a geriatric evaluation and management population. *J Rehabil Med* 2010; 42: 956–961.
  42. Johnston M, de Morton N, Harding K, Taylor N. Measuring mobility in patients living in the community with Parkinson disease. *NeuroRehabilitation* 2013; 32: 957–966.
  43. de Morton N, Davidson M, Keating JL. Reliability of the de Morton mobility index (DEMMI) in an older acute medical population. *Physiother Res Int* 2010; 16: 159–169.
  44. Rockwood K, Awalt E, Carver D, MacKnight C. Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. *J Gerontol. A Biol Sci Med Sci* 2000; 55: 70–73.
  45. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther* 1996; 76: 1109–1123.