European Review of
Aging and Physical Activity

REVIEW ARTICLE

Open Access

Recommended motor assessments based on psychometric properties in individuals with dementia: a systematic review

Sandra Trautwein[1]* , Philipp Maurus[1,2], Bettina Barisch-Fritz[1], Anela Hadzic[1] and Alexander Woll[1]

Abstract

Background: Motor assessments are important to determine effectiveness of physical activity in individuals with dementia (IWD). However, inappropriate and non-standardised assessments without sound psychometric properties have been used. This systematic review aims to examine psychometric properties of motor assessments in IWD combined with frequency of use and effect sizes and to provide recommendations based on observed findings. We performed a two-stage systematic literature search using Pubmed, Web of Science, Cochrane Library, ALOIS, and Scopus (inception - July/September 2018, English and German). The first search purposed to identify motor assessments used in randomised controlled trials assessing effectiveness of physical activity in IWD and to display their frequency of use and effect sizes. The second search focused on psychometric properties considering influence of severity and aetiology of dementia and cueing on test-retest reliability. Two reviewers independently extracted and analysed findings of eligible studies in a narrative synthesis.

Results: Literature searches identified 46 randomised controlled trials and 21 psychometric property studies. While insufficient information was available for validity, we observed sufficient inter-rater and relative test-retest reliability but unacceptable absolute test-retest reliability for most assessments. Combining these findings with frequency of use and effect sizes, we recommend Functional Reach Test, Groningen Meander Walking Test (time), Berg Balance Scale, Performance Oriented Mobility Assessment, Timed Up & Go Test, instrumented gait analysis (spatiotemporal parameters), Sit-to-Stand assessments (repetitions> 1), and 6-min walk test. It is important to consider that severity and aetiology of dementia and cueing influenced test-retest reliability of some assessments.

Conclusion: This review establishes an important foundation for future investigations. Sufficient relative reliability supports the conclusiveness of recommended assessments at group level, while unacceptable absolute reliability advices caution in assessing intra-individual changes. Moreover, influences on test-retest reliability suggest tailoring assessments and instructions to IWD and applying cueing only where it is inevitable. Considering heterogeneity of included studies and insufficient examination in various areas, these recommendations are not comprehensive. Further research, especially on validity and influences on test-retest reliability, as well as standardisation and development of tailored assessments for IWD is crucial.
This systematic review was registered in PROSPERO (CRD42018105399).

Keywords: Physical performance measurements, Cognitive impairment, Validity, Reliability, Frequency of use

* Correspondence: sandra.trautwein@kit.edu
[1]Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Engler-Bunte-Ring 15, 76131 Karlsruhe, Germany
Full list of author information is available at the end of the article

Background

Physical activity has gained importance as therapeutic strategy for individuals with dementia (IWD), and in accordance, the number of trials investigating its effectiveness on motor and cognitive performance in IWD has increased [1]. However, methodological limitations, such as inappropriate or inconclusive motor assessments, affect the derivation of evidence. Thus, further high quality investigations are required [2–4].

Considering motor assessments, high quality is reflected by appropriateness for the intended population, sensitivity to change, sound psychometric properties, and standardisation [4–6]. In many cases, motor assessments used in previous trials failed to meet these criteria. The majority of applied assessments has predominately been developed for healthy older adults and does not consider specific characteristics of IWD [7]. However, IWD and unimpaired individuals differ in their cognitive and motor performance [8–12]. Thus, tailoring motor assessments to IWD is essential to ensure appropriateness. Furthermore, insufficient or inconsistent research regarding sensitivity to change and psychometric properties in IWD [13] restricts the derivation of meaningful conclusions from applied motor assessments [14, 15]. Referring to this, literature indicates that dementia affects reliability [6, 16–18], which was scarcely considered in previous trials. With regard to standardisation, previous research utilised a variety of motor assessments and modifications, affecting comparability [4, 13]. Therefore, inappropriateness, insensitivity, inconclusiveness, and non-standardisation limit the derivation of evidence.

Considering heterogeneous cognitive and motor impairments [10, 19], motor assessments may not be equally suitable for all IWD. Severity and aetiology of dementia, which are important determinants contributing to this heterogeneity [19, 20], potentially influence psychometric properties of motor assessments. Particularly, test-retest reliability may decrease with increasing severity of dementia, due to growing intra-individual variability or progressive difficulties to participate in motor assessments [6, 16–18]. Similarly, aetiology of dementia can influence test-retest reliability as cognitive and motor impairments vary in time of occurrence and severity in different aetiologies [14, 19]. Moreover, the influence of external cues on test-retest reliability, which are used to compensate for cognitive and motor impairments, has been discussed [16, 21].

Literature comprehensively addressing motor assessments for IWD is limited. The importance of research in this area is highlighted in a qualitative approach [22] of analysing the appropriateness of motor assessments for IWD. Additionally to elaborating recommendations, this article emphasises the need for tailoring and standardising motor assessments for IWD [22]. Moreover, three systematic reviews [7, 13, 23] and one scoping review [24] examined frequency of use, sensitivity to change, and psychometric properties. Bossers et al. [13] and McGough et al. [24] identified eight frequently applied, sensitive assessments, showing good to excellent relative test-retest reliability. Fox et al. [7] found appropriate relative test-retest reliability, but insufficient absolute test-retest reliability and limited information on validity for several motor assessments. While Lee et al. [23] determined similar intraclass correlation coefficients (ICC), they applied a more stringent rating, suggesting acceptable relative test-retest reliability only for the Berg Balance Scale (BBS). Additionally, they considered the influence of different aetiologies of dementia on relative test-retest reliability, but were not able to draw conclusions due to insufficient research. In summary, these reviews provide an important basis, but do not actually allow a comprehensive quantitative evaluation of motor assessments for IWD. Previous reviews focused on frequency of use and sensitivity to change [13, 24] or just considered relative reliability and neglected other psychometric properties such as absolute reliability or validity [13, 23, 24]. They only investigated psychometric properties of the most common motor assessments without taking into account the influences of the heterogeneity of IWD [7, 13, 24] or considering further outcomes such as frequency of use or sensitivity to change [7, 23]. Moreover, information on how psychometric properties were graded was rare [13, 23, 24], no specific recommendations were suggested [7, 23], and the results of different outcomes were not combined when drawing conclusions [7]. Finally, previous randomised controlled trials (RCT) with IWD applied additional motor assessments which were not considered in previous reviews [7, 13, 23, 24].

With respect to these limitations, we indicated the following main research gaps: (a) comprehensive quantitative approaches combining outcomes of identified reviews including psychometric properties, frequency of use, and effect sizes of motor assessments applied in previous RCT with IWD and (b) research on the influence of severity and aetiology of dementia and cueing on test-retest reliability. Therefore, the objectives of this systematic review are: (1) to quantitatively examine motor assessments for IWD used in previous RCT by comprehensively analysing psychometric properties (primary outcome), frequency of use, and effect sizes of those assessments (secondary outcomes) and (2) to assess the influence of severity and aetiology of dementia and cueing on test-retest reliability. Based on primary and secondary outcomes, this review derives recommendations, which contribute to create consensus and decrease heterogeneity of motor assessments for future research. It needs to be considered that there are several purposes and reasons for applying motor assessments. Motor assessments are essential for diagnostic purposes and to assess changes over time, e.g. in RCT.

Regarding specific reasons, they are utilised to determine actual motor performance, but also to evaluate related outcomes, such as frailty [25] and risk of falls [26], or to draw conclusions on underlying cognitive performance [27]. This review focuses on motor assessments to assess changes over time, but does not further differentiate between various reasons for the use of motor assessments. Instead, it aims to provide a general overview.

Methods

For this systematic review, we considered the guidelines and recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement [28, 29]. Furthermore, we registered the systematic review in PROSPERO (CRD42018105399).

We performed a two-stage literature search to address the objectives of this systematic review. A first search focused on the identification of motor assessments applied in RCT in IWD. Based on these findings, a second search (main search) aimed to determine publications examining psychometric properties of the identified motor assessments. This approach ensures to focus on those motor assessments commonly applied in IWD and allows the determination of various outcomes required for a comprehensive quantitative evaluation of motor assessments for IWD. The taxonomy of COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative [30] provided the terminology and definitions of psychometric properties. In line with literature, we applied the terms relative and absolute reliability for reliability and measurement error, respectively [31]. Relative reliability, quantified by correlation coefficients, refers to the degree to which individual measurements maintain their position within a sample over repeated assessments, while absolute reliability, quantified by standard error of measurements or minimal detectable changes, is the degree to which individual measurements vary over repeated assessments [6, 31, 32].

First search

For the first search, we examined the electronic databases Pubmed, Web of Science, Cochrane Library, and ALOIS between December 2016 and July 2018 without date restrictions. We applied terms related to dementia, physical activity, and motor performance to identify eligible trials (see Additional file 1 for complete search term), supplemented by manually checking references of indicative articles and reviews. Two reviewers independently screened titles and abstracts (ST and BB) and checked inclusion criteria during full-text analysis (ST and AH). Trials were eligible if they met the following criteria: (a) designed as (cluster) RCT, (b) included individuals with primary dementia (Alzheimer's disease (AD), vascular dementia, frontotemporal dementia, and

Lewy body disease) older than 65 years, (c) applied physical activity interventions,[1] (d) used motor assessments independent of intended reasons, and (e) were published and written in English or German. We excluded comments, conference abstracts, protocols, and trial registrations. If there were disagreements, the two reviewers consulted a third reviewer (AW) to reach a consensual decision.

One reviewer (ST) extracted the following data from included RCT using a standardised extraction form: sample size, sample characteristics, motor assessments, means and standard deviations of baseline and post motor assessments, corresponding F/t statistics, and effect sizes. A second reviewer (AH) checked the outcomes. The two reviewers discussed ambiguities and disagreements in consensus meetings and consulted a third reviewer (BB) if they reached no agreement.

In addition to analysing frequency of use of identified motor assessments, we calculated time*group interaction effect sizes to represent their sensitivity to change. We determined Cohen's d if F (time*group interaction) or t (between group baseline-post differences) statistics, or baseline-post differences including standard deviations were provided ([34] formulas see Additional file 2). A Cohen's d of 0.2, 0.5, and 0.8 represents a small, medium, and large effect size, respectively [35]. Furthermore, we considered time*group interaction effect sizes provided in RCT.

This first search primarily aimed to identify motor assessments used in previous RCT with IWD and served as basis for the main search. Hence, we did not assess risk of bias.

Main search

For the main search, we examined the electronic databases PubMed, Web of Science, Cochrane Library, and Scopus (no date restrictions) between August and September 2018 for terms related to dementia, psychometric properties, and motor assessments identified in the first search (see Additional file 3 for complete search term). Additionally, we manually checked reference lists of indicative articles. Two reviewers (ST and PM) independently screened titles and abstracts and checked inclusion criteria during full-text analysis. Trials were eligible if they fulfilled the following criteria: (a) examined psychometric properties (content validity, construct validity, criterion validity, internal consistency, intra-rater reliability, inter-rater reliability, test-retest reliability, relative and absolute reliability) of (b) motor assessments in (c) individuals with primary dementia (AD, vascular dementia, frontotemporal

---

[1]defined as all types of physical activity that are planned, structured, repetitive, and purposive aiming to improve or maintain one or more components of physical fitness [33]

dementia, and Lewy body disease) aged above 65 years, (d) applied Mini-Mental State Examination (MMSE) [36], and (e) were written and published in English or German. We excluded comments and conference abstracts. The two reviewers discussed disagreements and consulted a third reviewer (BB) to resolve remaining discrepancies.

Two reviewers (ST and PM) independently extracted the following information from eligible investigations utilising a standardised data extraction form: sample size, sample characteristics, motor assessments, methodologies, and statistics of psychometric properties. Moreover, they independently assessed risk of bias of individual investigations with the COSMIN checklist [37, 38]. The two reviewers resolved disagreements through discussion and consulted a third reviewer (BB) if necessary.

Afterwards, we analysed findings of eligible investigations in a systematic narrative synthesis and summarised extracted information. In order to allow comparability of minimal detectable change values, we calculated percentage minimal detectable changes at 95% confidence interval ($MDC_{95\%}$) if any standard error of measurement or minimal detectable change was reported ([39, 40] formulas: see Additional file 4).

Moreover, we rated the results of each study against the COSMIN criteria for good measurement properties [41]. Since information on minimal important change of considered motor assessments in IWD is rare [17], and no other firm criteria for acceptable values [42] are available, we considered a $MDC_{95\%}$ higher than 30% as unacceptable [43, 44]. Based on COSMIN reliability criteria for good measurement properties [41] and indications for unacceptable values [43, 44], we rated relative and absolute reliability as follows:

- sufficient relative/absolute reliability (+): ICC ≥ 0.70/ minimal detectable change at 95% confidence interval < minimal important change
- indeterminate relative/absolute reliability (?): ICC not reported/minimal important change not defined
- insufficient relative/absolute reliability (−): ICC < 0.70/minimal detectable change at 95% confidence interval > minimal important change
- unacceptable absolute reliability (↓): $MDC_{95\%}$ > 30%

Subsequently, we summarised overall evidence and graded quality of evidence using the Grading of Recommendations Assessment, Development, and Evaluation approach, which considers risk of bias, inconsistency, imprecision, and indirectness of included investigations [41, 45]. Additionally, we analysed the influence of severity and aetiology of dementia and cueing on test-retest reliability. Therefore, we determined severity of dementia according to reported MMSE values (mild: MMSE = 26–17, moderate: MMSE = 17–10, severe: MMSE< 10

[46–48]) and/or classification of publications if range of MMSE was not reported. Due to insufficient information on aetiology, we were only able to compare between AD and various or not reported types. In accordance with Muir-Hunter et al. [14] we defined cueing as "providing any additional verbal, visual, or tactile direction necessary to ensure correct performance of the task after the initial set of standardized instructions was given". To investigate its influence on test-retest reliability, we classified cueing in five categories, considering information in identified psychometric property studies: (a) not reported, (b) no cueing, (c) verbal cueing, (d) verbal and visual/tactile cueing, and (e) more extensive cueing than (c) and (d) including physical assistance.

Results

Systematic searches (first and main search)

The first search revealed 5007 publications. After removing duplicates and initial screening on titles and abstracts, we screened the full texts of 309 publications and included 46 RCT for further analysis. For the main search, we obtained 902 publications. Removing duplicates and initial screening on titles and abstracts yielded 68 publications, of which we scanned full texts. Eventually, we included 21 eligible investigations in the narrative data synthesis (see Fig. 1, further information on study characteristics and data extractions are provided in Additional files 5, 6, 7 and 8).
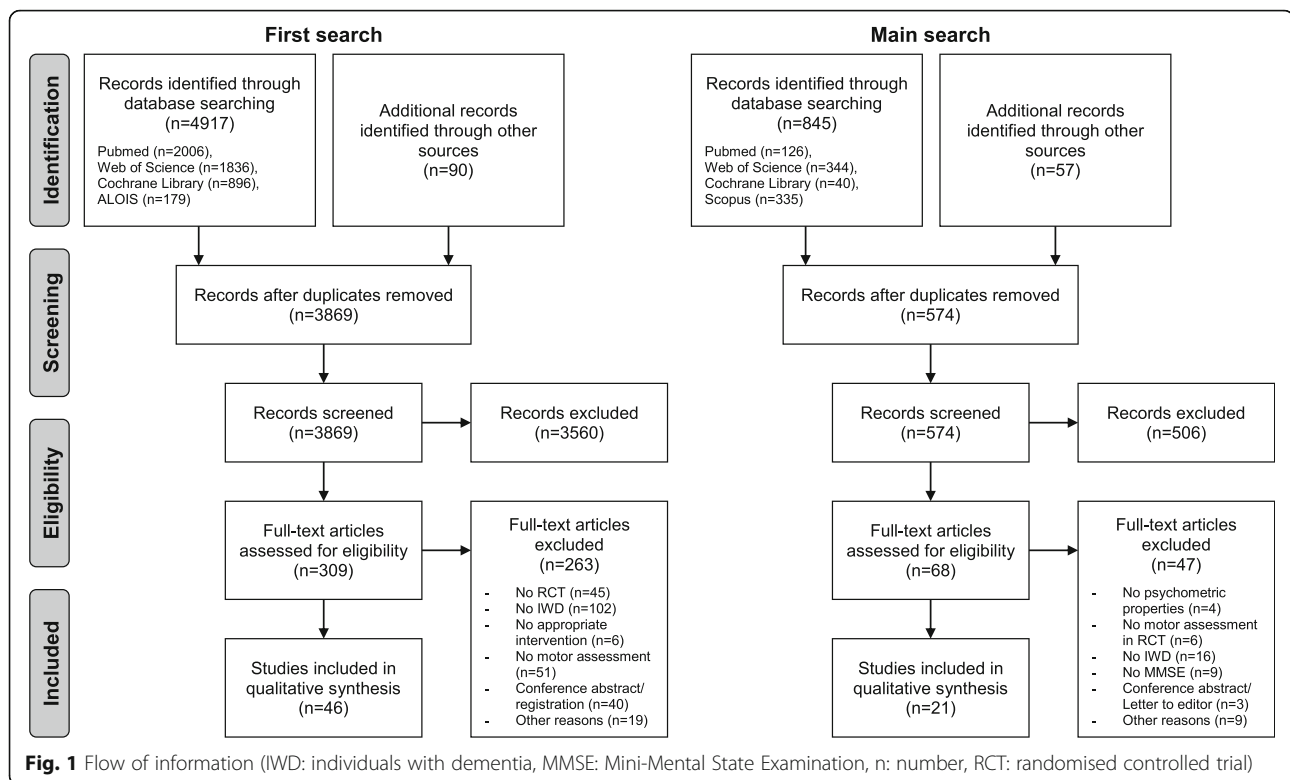
Motor assessments applied in previous randomised controlled trials

Previous RCT with IWD utilised 57 different motor assessments to determine balance, mobility and gait, strength, endurance, flexibility, and functional performance. Psychometric properties of 28 of these assessments were investigated in IWD. Table 1 contains a short description of all identified motor assessments with available psychometric property studies (see Additional file 9 for motor assessments identified during first search without available information on psychometric properties).

Psychometric properties

Seventeen of twenty-one studies examining psychometric properties focused on inter-rater and/or test-retest reliability. Herein, they determined consistency among different evaluators simultaneously rating the same participant, and between repeated measurements, respectively [32]. Investigations assessing content, construct, and criterion validity, internal consistency, and intrarater reliability were rare. Thus, we only summarised results and did not derive conclusions.

Fig. 1 Flow of information (IWD: individuals with dementia, MMSE: Mini-Mental State Examination, n: number, RCT: randomised controlled trial)

Summary for content, construct, and criterion validity, internal consistency, and intra-rater reliability[2]

The systematic search did not identify any investigation examining content validity. Based on hypotheses testing or revealing known group differences, construct validity was suggested for Physiomat assessments, the Erlangen Test of Activities of Daily Living (E-ADL Test), and knee extensor strength assessed with dynamometers [53, 110, 111, 114]. Seven investigations include information on criterion validity (concurrent and predictive validity), correlation with, or prediction of external criteria. For the E-ADL Test, criterion related validity was determined based on the relation between achieved scores and level of care [111]. Concurrent validity with spatiotemporal gait parameters or 2D-video motion analysis was established for a modified BBS, Short Physical Performance Battery (SPPB), and Assessment of Compensatory Sit-to-Stand Maneuvers in People With Dementia (ACSID) [26, 99]. Moreover, both the SPPB and 6-min walk test (6 min WT) significantly correlated with peak oxygen consumption (assessed with a cycle ergometer test), suggesting that these assessments are useful in identifying individuals with low aerobic capacity [115]. Furthermore, knee extensor strength was found to be a significant predictor for

several activities of daily living, gait, and sit-to-stand (STS) performance [114, 116]. No predictive validity concerning future falls could be observed for Timed Up & Go Test (TUG), Performance Oriented Mobility Assessment (POMA), and Five Times Sit-to-Stand Test (5x STS) [117].

Considering internal consistency, three studies observed Cronbach's α between 0.37 and 0.77 for E-ADL Test [110, 111] and 0.95 for BBS [15]. Furthermore, one study examining ACSID total score determined intra-rater reliability based on ICC ranging between 0.72 and 0.90 [99].

Inter-rater reliability (relative and absolute reliability)

Five studies assessed inter-rater reliability of nine assessments. ICC ranged from 0.72 to 1.00 and MDC_{95} included values between 0.0 and 98.0% [14, 15, 43, 99, 118]. Accordingly, all assessments reached sufficient relative inter-rater reliability. Quality of evidence for relative inter-rater reliability was high for BBS, moderate for TUG, and low or very low for all other assessments. Grading $MDC_{95\%}$, TUG and 6-m walk test (6 m WT) showed sufficient absolute inter-rater reliability, while it was insufficient/unacceptable for 4-m walk test (4 m WT), and indeterminate for all other assessments. Quality of evidence for absolute inter-rater reliability was low for 6 m WT and 30-s chair stand test (30s CST), and moderate for all remaining assessments (see Table 2).

---

[2]This summary utilises psychometric property terms indicated in original studies. These terms have not been consistently used throughout the literature and should have been adapted according to the COSMIN checklist [41].

Table 1 Description, frequency of use, and effect sizes of motor assessments applied in previous randomised controlled trials

Motor assessment	Description	Frequency of use	Time*group interaction effect size
Balance			
FICSIT-4 [49]	*Task*: performing four different stances with eyes open for 10 seconds: (a) feet together, (b) semi-tandem, (c) tandem, (d) single-leg *Measurement*: score [0–5], which rates performance according to ability to maintain stances	1 RCT (*n* = 109) [50]	–
Modified Clinical Test of Sensory Interaction of Balance [51]	*Task*: standing on a platform (NeuroCom Balance Master) as quietly as possible for 10 seconds under four sensory conditions: eyes open and closed standing on firm surface and foam *Measurements*: sway velocity [deg/s], composite score for all conditions	1 RCT (*n* = 40) [52]	–
Limits of Stability [51]	*Task*: standing on NeuroCom Balance Master and moving cursor from centre box directly to eight target boxes as fast and as close as possible by shifting weight *Measurements*: reaction time [s], movement velocity [deg/s], maximum excursion [%], directional control [%], summary composite score	1 RCT (*n* = 40) [52]	–
Physiomat-Trail-Making Task [53]	*Task*: standing on Physiomat and connecting digits by shifting weight *Measurements*: total duration [s], accuracy of sway path [digits/ms]	1 RCT (*n* = 84) [54]	–
Physiomat-Follow-The-Ball Task [53]	*Task*: standing on Physiomat and moving cursor from centre of screen directly to targets as fast as possible by shifting weight *Measurements*: total duration [s], accuracy of sway path [digits/ms]	1 RCT (n = 84) [54]	–
FR [55]	*Task*: standing next to a wall, holding one arm parallel to a metre stick attached to the wall at shoulder height, and reaching forward as far as possible without losing balance or changing foot position *Measurement*: distance from starting to end position [cm]	5 RCT (*n* = 204) [52, 56–59]	Small to large [c]
Hill Step Test [60]	*Task*: stepping 1 foot onto a block and returning it to the floor as quickly as possible for 15 seconds *Measurement*: number of repetitions	2 RCT (*n* = 54) [52, 61]	–
Step Quick Turn Test [51]	*Task*: taking two steps forward on NeuroCom Balance Master, quickly turning, and returning to starting point *Measurements*: turn time [s], turn sway [deg/s]	1 RCT (n = 40) [52]	–
Figure of Eight Test [62]	*Task*: walking a lap of a standard figure-eight trajectory as quickly and accurately as possible *Measurements*: walking speed [m/s], number of oversteps	1 RCT (n = 109) [50]	–
GMWT [63]	*Task*: walking over a meandering curved line as quickly and accurately as possible *Measurements*: walking speed [m/s], number of oversteps	1 RCT (*n* = 109) [50]	–
BBS [64]	*Task*: 14-item functional balance assessment with simple everyday tasks (reaching, bending, transferring, standing, and rising), which are graded on a five-point ordinal scale (0 to 4) *Measurement*: score [0–56]	11 RCT (*n* = 648) [56, 57, 65–73]	Small to large [c/r]
Modified BBS [64]	*Task*: abbreviated version of the original 14-item BBS, excluding three items (chair-to-chair transfer, forward reach with outstretched arm, and alternate	1 RCT (*n* = 23) [74]	–

Table 1 Description, frequency of use, and effect sizes of motor assessments applied in previous randomised controlled trials *(Continued)*

Motor assessment	Description	Frequency of use	Time*group interaction effect size
	stepping on-off stool) *Measurement*: score [0–44]		
POMA [75]	*Task*: scale with two parts, assessing balance (B) and gait (G) (B) sitting balance, rising from a chair and sitting down, standing balance (with eyes open and closed), and turning balance (G) gait initiation, step length and height, symmetry, continuity, path direction, and trunk sway *Measurements*: total score [0–28], balance score [0–16], gait score [0–12]	7 RCT (*n* = 300) [68, 70, 76–80]	No to large [c/r]
Mobility and gait			
TUG [81]	*Task*: standing up from a chair, walking three metres, turning around, walking back to chair, and sitting down *Measurements*: time [s], number of steps	16 RCT (*n* = 1001) [50, 52, 56, 58, 59, 66, 68, 70, 73, 78, 79, 82–86]	No to large [c/r]
Cognitive TUG [87]	*Task*: TUG with additional cognitive task (counting backwards by threes/evoke names of animals) *Measurement*: time [s]	2 RCT (*n* = 60) [52, 56]	–
Manual TUG [25, 87]	*Task*: TUG with additional manual task (carrying a glass of water) *Measurement*: time [s]	1 RCT (n = 40) [52]	–
6 m WT [88]	*Task*: walking six metres with comfortable pace *Measurements*: walking speed [m/s], step length [m]	3 RCT (*n* = 379) [50, 71, 89]	–
4 m WT [88]	*Task*: walking four metres with comfortable pace *Measurement*: walking speed [m/s]	2 RCT (*n* = 244) [90, 91]	Small [c/r]
Instrumented gait analysis [92]	*Task*: walking with comfortable/fast pace over an electronic walkway (GAITRite, Bessou locometer, NeuroCom Balance Master) *Measurements*: walking speed [cm/s, m/s], cadence [steps/min], stride/step length [cm, m], stride time [s], double/single support [% of stride time], double limb support time [s], step width [cm], step time variability [CV], Walk-Ratio [step length/cadence]	6 RCT (*n* = 370) [52, 78, 93–96]	Small to large [c/r]
Strength			
5x STS [97]	*Task*: performing five repetitions of the STS task without upper extremity assistance *Measurement*: time [s]	7 RCT (*n* = 358) [52, 58, 68, 77, 78, 94, 98]	No to large [c/r]
STS on NeuroCom Balance Master [51]	*Task*: standing up from a seated position without upper extremity assistance *Measurements*: rising index [% of body weight], centre of gravity sway velocity [deg/s]	1 RCT (*n* = 40) [52]	–
ACSID [99]	*Task*: performing five repetitions of the STS task without upper extremity assistance while motor and cognitive aspects of movement process are qualitatively rated *Measurements*: total score [0–10], sub scores 'recall and initiation' [0–5], 'effective performance' [0–5]	1 RCT (*n* = 77) [100]	Large [c/r]
30s CST [17, 101]	*Task*: performing as many repetitions of STS task as possible in 30 s *Modified version*: use of upper extremity assistance is allowed *Measurement*: number of repetitions	5 RCT (*n* = 408) [56, 71, 74, 80, 84] Modified: 1 RCT (*n* = 109) [50]	Large [c/r]
Handgrip dynamometer [102]	*Task*: putting maximum force on a dynamometer	3 RCT (*n* = 263) [67, 78, 94]	No [r]

Table 1 Description, frequency of use, and effect sizes of motor assessments applied in previous randomised controlled trials
(Continued)

Motor assessment	Description	Frequency of use	Time*group interaction effect size
Maximum isometric strength assessed with dynamometers [103]	*Measurement*: maximum handgrip strength [KPa, kg] *Task*: pushing as hard as possible against a dynamometer after adopting a standardised position *Measurements*: maximum strength [N] and integral over time [Ns] for knee extension, knee flexion, and ankle flexion	2 RCT (*n* = 216) [50, 78]	–
Endurance			
6 min WT [104]	*Task*: walking for 6 minutes with comfortable pace *Measurement*: distance [m, ft]	5 RCT (*n* = 359) [50, 57, 105–107]	–
Functional performance			
SPPB [108]	*Task*: three subtests including standing balance (tandem, semi-tandem, and side-by-side stands), walking speed over an 8-ft walking course, and 5x STS *Measurement*: score [0–12]	3 RCT (*n* = 313) [77, 90, 109]	Small to medium [c/r]
E-ADL Test [110, 111]	*Task*: five items (pouring a drink, spreading butter on a sandwich and cutting the sandwich, open a small cupboard with a key, washing and drying hands, and tying a bow on a small wrapped present), which are rated according to correctly performed substeps (0–6 points) *Measurement*: score [0–30]	2 RCT (*n* = 192) [112, 113]	–

4 m WT 4-m walk test, *5x STS* Five Times Sit-to-Stand Test, *6 m WT* 6-m walk test, *6 min WT* 6-min walk test, *30s CST* 30-s chair stand test, *ACSID* Assessment of Compensatory Sit-to-Stand Maneuvers in People With Dementia, *BBS* Berg Balance Scale, *E-ADL Test* Erlangen Test of Activities of Daily Living, *FICSIT-4* Frailty and Injuries: Cooperative Studies of Intervention Techniques - subtest 4, *FR* Functional Reach Test, *GMWT* Groningen Meander Walking Test, *n* Number of analysed participants, *POMA* Performance Oriented Mobility Assessment, *RCT* Randomised controlled trial/s, *SPPB* Short Physical Performance Battery, *STS* Sit-to-Stand, *TUG* Timed Up & Go Test.
[c] calculated effect size, [r] effect size provided of randomised controlled trial

Regarding balance assessments, ICC were higher for Groningen Meander Walking Test (GMWT) and BBS than for Functional Reach Test (FR). Furthermore, $MDC_{95\%}$ were lower for BBS compared to GMWT. Focusing on GMWT, time measurement showed lower $MDC_{95\%}$ than number of oversteps. For mobility and gait, ICC increased and $MDC_{95\%}$ decreased from 4 m WT, through 6 m WT, to TUG. Considering strength assessments, ICC were higher for 30s CST counting repetitions than for ACSID rating STS performance, while $MDC_{95\%}$ was only determined for 30s CST. Since ICC was only assessed for 6 min WT, a comparison of inter-rater reliability of endurance assessments was not possible (see Table 2).

Test-retest reliability (relative and absolute reliability)
Fifteen studies investigated test-retest reliability considering 24 assessments. ICC ranged between 0.02 and 0.99 and $MDC_{95\%}$ varied from 6.8 to 225.7% [5, 6, 14, 17, 26, 43, 51, 53, 63, 102, 110, 114, 118, 120, 121] (see Table 3).

Most studies focused on between-day test-retest reliability, while some studies examined within-day and within-session test-retest reliability. Comparing these studies, ICC increased and $MDC_{95\%}$ decreased, respectively, from between-day (ICC = 0.02–0.99, $MDC_{95\%}$ = 6.8–225.7% [5, 14, 17, 43, 51, 53, 63, 102, 118, 120, 121]), through within-day (ICC = 0.79–0.99, $MDC_{95\%}$ = 21.1–30.0% [6, 26, 118]), to within-session test-retest reliability (ICC = 0.95–0.98 [114]).

Balance Six investigations assessing test-retest reliability of eleven balance assessments determined ICC and $MDC_{95\%}$ ranging between 0.32–0.99 and 10.2–225.7%, respectively [14, 17, 43, 51, 53, 63]. Relative test-retest reliability was sufficient for all balance assessments except for Limits of Stability, Step Quick Turn Test, and simple condition of Physiomat-Trail-Making Task. However, quality of evidence for relative test-retest reliability was low or very low for most assessments. Only GMWT (time) and BBS reached moderate quality of evidence. Absolute test-retest reliability for balance assessments was indeterminate or unacceptable with moderate to very low quality of evidence (see Table 3).

Table 2 Relative and absolute inter-rater reliability

Variable		Study	Relative inter-rater reliability							Absolute inter-rater reliability						
			ICC	Rating	Risk of bias	Inconsistency	Imprecision	Indirectness	Quality of evidence	MDC$_{95\%}$	Rating	Risk of bias	Inconsistency	Imprecision	Indirectness	Quality of evidence
Balance																
FR	Distance	1 study of adequate quality (n = 15) [14]	0.79	+	Serious	No	n < 50	No	Very low	Not assessed						
GMWT	Time	1 study of adequate/ very good quality (n = 53) [43]	0.99	+	Serious	No	n = 50–100	No	Low	14.5%	?	No	No	n = 50–100	No	Moderate
	Number of oversteps	1 study of adequate/ very good quality (n = 53) [43]	0.99	+	Serious	No	n = 50–100	No	Low	17.1%	?	No	No	n = 50–100	No	Moderate
BBS	Score	3/2 studies of adequate/very good quality (n = 101/86) [14, 15, 43]	0.72–0.99	+	No	No	No	No	High	5.9–7.1%	?	No	No	n = 50–100	No	Moderate
Mobility and gait																
TUG	Time	2 studies/1 study of adequate/very good quality (n = 68/53) [14, 43]	0.98–0.99	+	No	No	n = 50–100	No	Moderate	7.9%	+^{b}	No	No	n = 50–100	No	Moderate
6 m WT	Walking speed	1 study of adequate/ very good quality (n = 33) [15]	0.97	+	Serious	No	n < 50	No	Very low	15.7%	+^{c}	No	No	n < 50	No	Low
4 m WT	Time	1 study of adequate/ very good quality (n = 53) [43]	0.82	+	Serious	No	n = 50–100	No	Low	98.0%	-^{c}/↓	No	No	n = 50–100	No	Moderate
Strength																
ACSID	Score	1 study of very good quality (n = 94) [99]	0.85	+	No	No	n = 50–100	No	Moderate	Not assessed						
30s CST	Repetitions	1 study of adequate/ very good quality (n = 33) [15]	1.00	+	Serious	No	n < 50	No	Very low	0.0%	?	No	No	n < 50	No	Low
Endurance																
6 min WT	Distance	1 study of adequate quality (n = 33)^{a} [118]	0.97–0.99	+	Serious	No	n < 50	No	Very low	Not assessed						
	Walking speed	1 study of adequate quality (n = 33)^{a} [118]	0.96–0.98	+	Serious	No	n < 50	No	Very low	Not assessed						

4 m WT 4-m walk test, *6 m WT* 6-m walk test, *6 min WT* 6-min walk test, *30s CST* 30-s chair stand test, *ACSID* Assessment of Compensatory Sit-to-Stand Maneuvers in People With Dementia, *BBS* Berg Balance Scale, *FR* Functional Reach Test, *GMWT* Groningen Meander Walking Test, *ICC* Intraclass correlation coefficient, *MDC$_{95\%}$* Percentage minimal detectable changes at 95% confidence interval, *n* Total number of participants, *TUG* Timed Up & Go Test.

Rating according to COSMIN criteria for good measurement properties: + = sufficient, − = insufficient, ? = indeterminate, ↓ = unacceptable absolute inter-rater reliability.

^{a} inter-rater reliability was determined on 2 times of measurement, ^{b} minimal important change (TUG) = 10.1 s [17, 119], ^{c} minimal important change (walking speed) = 0.21 m/s [17, 119]

Table 3 Relative and absolute test-retest reliability

Variable	Study	Relative test-retest reliability		Risk of bias	Inconsistency	Imprecision	Indirectness	Quality of evidence
		ICC	Rating					
Balance								
FICSIT-4 — Score	1 study of adequate quality (n = 58)[a] [17]	0.79–0.82	+	Serious	No	n = 50-100	No	Low
Modified Clinical Test of Sensory Interaction of Balance — Sway velocity	1 study of adequate quality (n = 14) [51]	0.91	+	Serious	No	n < 50	No	Very low
Limits of Stability — Reaction time	1 study of adequate quality (n = 14) [51]	0.52	–	Serious	No	n < 50	No	Very low
Movement velocity	1 study of adequate quality (n = 14) [51]	0.48	–	Serious	No	n < 50	No	Very low
Maximum excursion	1 study of adequate quality (n = 14) [51]	0.68	–	Serious	No	n < 50	No	Very low
Directional control	1 study of adequate quality (n = 14) [51]	0.71	+	Serious	No	n < 50	No	Very low
Physiomat-Trail-Making Task [53] — Score	1 study of adequate quality (n = 74) [53]	0.90	+	Serious	No	n = 50-100	No	Low
Sway Path	1 study of adequate quality (n = 47–73)[b] [53]	0.47–0.82	+/– depending on condition	Serious	No	n = 50-100	No	Low
Time	1 study of adequate quality (n = 47–73)[b] [53]	0.55–0.83	+/– depending on condition	Serious	No	n = 50-100	No	Low
Physiomat-Follow-The-Ball Task — Sway Path	1 study of adequate quality (n = 73) [53]	0.84	+	Serious	No	n = 50-100	No	Low
Time	1 study of adequate quality (n = 73) [53]	0.79	+	Serious	No	n = 50-100	No	Low
FR — Distance	2 studies of adequate quality (n = 29) [14, 51]	0.81–0.84	+	No	No	n < 50	No	Low
Hill Step Test — Number of steps	1 study of adequate quality (n = 14) [51]	0.87	+	Serious	No	n < 50	No	Very low
Step Quick Turn Test — Time	1 study of adequate quality (n = 14) [51]	0.55	–	Serious	No	n < 50	No	Very low
Sway	1 study of adequate quality (n = 14) [51]	0.64	–	Serious	No	n < 50	No	Very low
Figure of Eight Test — Time	1 study of adequate quality (n = 46)[a] [17]	0.85–0.94	+	Serious	No	n < 50	No	Very low
GMWT — Time	2 studies of adequate quality (n = 95)[a] [43, 63]	0.93–0.99	+	No	No	n = 50-100	No	Moderate
Number of oversteps	2 studies of adequate quality (n = 95)[a] [43, 63]	0.57–0.96	?	No	Yes	n = 50-100	No	Not assigned (inconsistency)
BBS — Score	2 studies of adequate quality (n = 68) [14, 43]	0.95–0.99	+	No	No	n = 50-100	No	Moderate
Mobility and gait								
TUG — Time	6/5 studies of adequate quality (n = 200/191)[a] [6, 14, 17, 43, 51, 102]	0.72–0.99	+	No	No	No	No	High

Table 3 Relative and absolute test-retest reliability *(Continued)*

	Variable	Study	Relative test-retest reliability		Risk of bias	Inconsistency	Imprecision	Indirectness	Quality of evidence
			ICC	Rating					
Cognitive TUG	Time	1 study of adequate quality (n = 10) [51]	0.51	–	Serious	No	n < 50	No	Very low
Manual TUG	Time	1 study of adequate quality (n = 14) [51]	0.70	+	Serious	No	n < 50	No	Very low
6 m WT	Walking speed	1 study of adequate quality (n = 58)[a] [17]	0.83–0.89	+	Serious	No	n = 50–100	No	Low
	Time	1 study of adequate quality (n = 9–10)[b] [102]	0.92–0.95	+	Serious	No	n < 50	No	Very low
	Number of steps	1 study of adequate quality (n = 9–10)[b] [102]	0.80–0.90	+	Serious	No	n < 50	No	Very low
4 m WT	Time	1 study of adequate quality (n = 53) [43]	0.85	+	Serious	No	n = 50–100	No	Low
Instrumented gait analysis	Walking speed	4/3 studies of adequate quality (n = 93/85)[a, d, e] [6, 26, 51, 121]	0.50–0.98	+ (except for NeuroCom Balance Master)	No	No	n = 50–100	No	Moderate
	Step length	2 studies of adequate quality (n = 34)[a, d, e] [51, 121]	0.75–0.98	+	No	No	n < 50	No	Low
	Step width	2 studies of adequate quality (n = 34)[a, d, e] [51, 121]	0.89–0.95	+	No	No	n < 50	No	Low
	Stride length	2 studies/1 study of adequate quality (n = 28/20)[e] [26, 121]	0.97–0.98	+	No	No	n < 50	No	Low
	Cadence	2 studies/1 study of adequate quality (n = 28/20)[e] [26, 121]	0.88–0.91	+	No	No	n < 50	No	Low
	Swing time	2 studies/1 study of adequate quality (n = 28/20)[e] [26, 121]	0.89–0.96	+	No	No	n < 50	No	Low
	Stance time	1 study of adequate quality (n = 20)[e] [121]	0.70–0.73	+	Serious	No	n < 50	No	Very Low
	Toe in/out angle	1 study of adequate quality (n = 20)[e] [121]	0.91–0.93	+	Serious	No	n < 50	No	Very Low
	Walking speed variability	1 study of adequate quality (n = 16) [5]	0.66	–	Serious	No	n < 50	No	Very Low
	Stride length variability	1 study of adequate quality (n = 16) [5]	0.80	+	Serious	No	n < 50	No	Very Low
	Stride width variability	1 study of adequate quality (n = 16) [5]	0.83	+	Serious	No	n < 50	No	Very Low
	Cadence variability	1 study of adequate quality (n = 16) [5]	0.65	–	Serious	No	n < 50	No	Very Low
Strength									
5x STS	Time	2 studies/1 study of adequate quality (n = 24/14) [51, 102]	0.80–0.94	+	No	No	n < 50	No	Low

Table 3 Relative and absolute test-retest reliability *(Continued)*

	Variable	Study	Relative test-retest reliability		Risk of bias	Inconsistency	Imprecision	Indirectness	Quality of evidence
			ICC	Rating					
STS on NeuroCom Balance Master	Rising Index	1 study of adequate quality (n = 14) [51]	0.95	+	Serious	No	n < 50	No	Very low
	COG sway velocity	1 study of adequate quality (n = 14) [51]	0.02	–	Serious	No	n < 50	No	Very low
Modified 30s CST	Repetitions	1 study of adequate quality (n = 52)[a] [17]	0.79–0.88	+	Serious	No	n = 50–100	No	Low
Handgrip dynamometer	Force	3 studies/1 study of adequate quality (n = 143/57)[a] [17, 102, 120]	0.42–0.98	+ (except for severe dementia)	No	No	No	No	High
Maximum isometric strength assessed with dynamometers	Peak force	1 studies of adequate quality (n = 11–12)[f] [102]	0.63–0.71	?	Serious	Yes	n < 50	No	Not assigned (inconsistency)
	(Normalised) torque	1 studies of adequate quality (n = 60)[a] [114]	0.95–0.98	+	Serious	No	n = 50–100	No	Low
Endurance									
6 min WT	Distance	2 studies/1 study of adequate quality (n = 84/51)[a, c] [6, 118]	0.76–0.98	+	No	No	n = 50–100	No	Moderate
	Walking speed	1 study of adequate quality (n = 33)[c] [118]	0.75–0.89	+	Serious	No	n < 50	No	Very Low
Functional performance									
E-ADL Test	Score	1 study of doubtful quality (n = 42) [110]	r = 0.73[g]	?	Very serious	No	n < 50	No	Very Low

4 m WT 4-m walk test, *5x STS* Five Times Sit-to-Stand Test, *6 m WT* 6-m walk test, *6 min WT* 6-min walk test, *30s CST* 30-s chair stand test, *BBS* Berg Balance Scale, *COG* Centre of gravity, *E-ADL Test* Erlangen Test of Activities of Daily Living, *FICSIT-4* Frailty and Injuries: Cooperative Studies of Intervention Techniques - subtest 4, *FR* Functional Reach Test, *GMWT* Groningen Meander Walking Test, *ICC* Intraclass correlation coefficient, *MDC95%* Percentage minimal detectable changes at 95% confidence interval, *n* Total number of participants, *STS* Sit-to-Stand, *TUG* Timed Up & Go Test.
Rating according to COSMIN criteria for good measurement properties: + = sufficient, – = insufficient,? = indeterminate, ↓ = unacceptable absolute test-retest reliability.
[a] test-retest reliability was assessed for different subgroups, [b] test-retest reliability was assessed for different conditions, [c] test-retest reliability was assessed for 2 different raters and 2 different between-test intervals, [d] test-retest reliability was assessed with 2 different devices, [e] test-retest reliability was assessed with 2 analysis sets, [f] test-retest reliability was assessed for 3 muscle groups, [g] Spearman's rank correlation coefficient, [h] minimal important change (TUG) = 10.1 s [17, 119], [i] minimal important change (walking speed) = 0.21 m/s [17, 119]

Table 3 Relative and absolute test-retest reliability *(Continued)*

	Absolute test-retest reliability						
	MDC$_{95\%}$	Rating	Risk of bias	Inconsistency	Imprecision	Indirectness	Quality of evidence
Balance							
FICSIT-4	58.9–71.1%	→	Serious	No	n = 50–100	No	Low
Modified Clinical Test of Sensory Interaction of Balance	36.5%	→	Serious	No	n < 50	No	Very low
Limits of Stability	38.0%	→	Serious	No	n < 50	No	Very low
	38.9%	→	Serious	No	n < 50	No	Very low
	15.9%	?	Serious	No	n < 50	No	Very low
	21.8%	?	Serious	No	n < 50	No	Very low
Physiomat-Trail-Making Task [53]	Not assessed						
	Not assessed						
	Not assessed						
Physiomat-Follow-The-Ball Task	Not assessed						
	Not assessed						
FR	15.4–68.9%	?/↓	No	Yes	n < 50	No	Not assigned (inconsistency)
Hill Step Test	26.2%	?	Serious	No	n < 50	No	Very low
Step Quick Turn Test	38.1%	→	Serious	No	n < 50	No	Very low
	29.7%	?	Serious	No	n < 50	No	Very low
Figure of Eight Test	36.9–37.9%	→	Serious	No	n < 50	No	Very low
GMWT	19.6–31.2%	?/↓	No	No	n = 50–100	No	Moderate
	33.3–225.7%	→	No	Yes	n = 50–100	No	Not assigned (inconsistency)
BBS	10.2–38.6%	?/↓	No	No	n = 50–100	No	Moderate
Mobility and gait							
TUG	15.8–39.6%	+^{h}/↓	No	No	No	No	High
Cognitive TUG	36.2%	+^{h}/↓	Serious	No	n < 50	No	Very low
Manual TUG	26.7%	+^{h}	Serious	No	n < 50	No	Very low
6 m WT	31.6–41.5%	-^{i}/↓	Serious	No	n = 50–100	No	Low
	Not assessed						
	Not assessed						
4 m WT	84.3%	-^{i}/↓	Serious	No	n = 50–100	No	Low
Instrumented gait analysis	10.2–48.3%	+^{i}/↓	No	No	n = 50–100	No	Moderate
	7.0–35.6%	?/↓	No	No	n < 50	No	Low
	20.0–24.7%	?	No	No	n < 50	No	Low
	6.8–8.5%	?	Serious	No	n < 50	No	Very low

Table 3 Relative and absolute test-retest reliability *(Continued)*

	Absolute test-retest reliability						
	MDC$_{95\%}$	Rating	Risk of bias	Inconsistency	Imprecision	Indirectness	Quality of evidence
	7.1–7.5%	?	Serious	No	n < 50	No	Very low
	7.0–7.1%	?	Serious	No	n < 50	No	Very low
	8.6–8.7%	?	Serious	No	n < 50	No	Very low
	28.2–33.5%	?/↓	Serious	No	n < 50	No	Very low
	77.8%	→	Serious	No	n < 50	No	Very low
	71.7%	→	Serious	No	n < 50	No	Very low
	46.9%	→	Serious	No	n < 50	No	Very low
	41.4%	→	Serious	No	n < 50	No	Very low
Strength							
5x STS	29.9%	?	Serious	No	n < 50	No	Very low
STS on NeuroCom Balance Master	21.8%	?	Serious	No	n < 50	No	Very low
	80.2%	→	Serious	No	n < 50	No	Very low
Modified 30s CST	33.2–45.7%	→	Serious	No	$n = 50–100$	No	Low
Handgrip dynamometer	34.9–36.8%	→	Serious	No	$n = 50–100$	No	Low
Maximum isometric strength assessed with dynamometers	Not assessed						
	Not assessed						
Endurance							
6 min WT	21.2–28.9%	?	Serious	No	$n = 50–100$	No	Low
	Not assessed						
Functional performance							
E-ADL Test	Not assessed						

GMWT (time) and BBS showed the highest ICC, while we could not observed a clear tendency for MDC$_{95\%}$. Comparing different outcomes of GMWT, ICC were higher and MDC$_{95\%}$ were lower for time than for number of oversteps (see Table 3).

Mobility and gait Nine studies investigated test-retest reliability of six mobility and gait assessments. They reported ICC between 0.50 and 0.99 and MDC$_{95\%}$ from 6.8 to 84.3% [5, 6, 14, 17, 26, 43, 51, 102, 121]. Relative test-retest reliability was sufficient for TUG, manual TUG, 6 m WT, 4 m WT, and instrumented gait analysis (except for cadence variability, walking speed variability, and walking speed assessed with NeuroCom Balance Master), while it was insufficient for cognitive TUG. Quality of evidence for relative test-retest reliability was high for TUG, moderate to very low for instrumented gait analysis, and low or very low for all other assessments. Absolute test-retest reliability was indeterminate for spatiotemporal gait parameters, insufficient/unacceptable for variability gait parameters, 4 m WT, and 6 m WT, and sufficient for manual TUG. For TUG, cognitive TUG, and walking speed assessed with instrumented gait analysis, absolute test-retest reliability was sufficient according to COSMIN criteria but unacceptable when applying MDC$_{95\%}$ limit of 30%. Except for TUG and walking speed assessed with instrumented gait analysis (high/moderate quality of evidence), quality of evidence for absolute test-retest reliability was low or very low (see Table 3).

Considering up and go tasks, ICC were higher for single than for dual task conditions. Focusing on short distance walk tests (WT), MDC$_{95\%}$ were lower for 6 m WT than for 4 m WT. Furthermore, the comparison of different gait parameters assessed with instrumented gait analysis, determined lower ICC and higher MDC$_{95\%}$ for variability measures than for spatiotemporal gait parameters. Comparing different assessments to determine short distance walking speed showed higher ICC and lower MDC$_{95\%}$ for instrumented gait analysis (except for NeuroCom Balance Master) than for simple short distance WT (see Table 3).

Strength Five studies focusing on test-retest reliability of strength assessments reported ICC and MDC$_{95\%}$ ranging between 0.02–0.98 and 21.8–80.2%, respectively [17, 51, 102, 114, 120]. Relative test-retest reliability was sufficient for modified 30s CST, 5x STS, handgrip dynamometers (except for severe dementia and one-time measuring), and maximum isometric strength assessed with dynamometers (except for dorsiflexor and iliopsoas muscle strength), while it was insufficient for STS on NeuroCom Balance Master (except for Rising Index). Quality of evidence for relative test-retest reliability was

high for handgrip dynamometers and low or very low for all other strength assessments. Absolute test-retest reliability was indeterminate for 5x STS and Rising Index of STS on NeuroCom Balance Master, and unacceptable for modified 30s CST, centre of gravity sway velocity of STS on NeuroCom Balance Master, and handgrip dynamometers. Quality of evidence for absolute test-retest reliability was low or very low for all assessments (see Table 3).

Comparing different STS assessments, ICC for assessments performing only one STS repetition were lower (except for Rising Index) than STS assessments with more repetitions. Moreover, MDC$_{95\%}$ increased from 5x STS, through modified 30s CST, to STS on NeuroCom Balance Master (except for Rising Index) (see Table 3).

Endurance Considering endurance, test-retest reliability was only determined for 6 min WT. Two studies observed ICC between 0.75 and 0.98, while MDC$_{95\%}$ ranged from 21.2 to 28.9% [6, 118]. Accordingly, relative test-retest reliability was sufficient with moderate to very low quality of evidence. Absolute test-retest reliability was indeterminate with low quality of evidence (see Table 3).

Functional performance Functional performance was rarely assessed. One study focusing on the E-ADL Test did not determine ICC and MDC$_{95\%}$, but found significant correlations for the whole test (r = 0.73) and separate items (r = 0.35–0.63) [110]. Quality of evidence was very low.

Influence of severity and aetiology of dementia and cueing on test-retest reliability

With respect to severity of dementia, the Frailty and Injuries: Cooperative Studies of Intervention Techniques - subtest 4 (FICSIT-4) and GMWT tend to yield higher ICC and/or lower MDC$_{95\%}$ with less cognitive impairment. In contrast, ICC were slightly higher and/or MDC$_{95\%}$ lower with stronger cognitive impairment for BBS, 6 m WT, modified 30s CST, and 5x STS (see Table 4).

Regarding aetiology of dementia, maximum isometric strength assessed with dynamometers and short distance walking speed (except for instrumented gait analysis with NeuroCom Balance Master) resulted in somewhat higher ICC and/or lower MDC$_{95\%}$ for AD vs. various or not reported types. In contrast, ICC were slightly higher and/or MDC$_{95\%}$ were lower for various or not reported types vs. AD for BBS, TUG (between-day reliability), up and go tasks in general (between-day reliability), 5x STS, and STS tasks in general (except for Rising Index) (see Table 5).

Considering cueing, GMWT and TUG showed somewhat higher ICC and/or lower MDC$_{95\%}$ when cueing was allowed or more extensive. In contrast, ICC were slightly higher and/or MDC$_{95\%}$ were lower for no cueing or less extensive cueing in FR, short distance WT, and short distance walking speed (see Table 6).

Table 4 Subgroup analysis of test-retest reliability considering severity of dementia

	Mild dementia	Mild to moderate dementia	Moderate dementia	Severity not reported
FICSIT-4	MMSE [mean (SD)]: 22.7 (2.1) ICC = 0.82 MDC$_{95\%}$ = 58.9% [17]	MMSE [mean (SD)]: 19.2 (4.4) ICC = 0.79 MDC$_{95\%}$ = 59.4% [17]	MMSE [mean (SD)]: 15.5 (2.4) ICC = 0.80 MDC$_{95\%}$ = 71.1% [17]	
GMWT	MMSE [mean (SD)]: n.r. ICC = 0.79–0.96 MDC$_{95\%}$ = n.r [63].	MMSE [mean (SD)]: 17.4 (4.3) ICC = 0.63–0.94 MDC$_{95\%}$ = 31.2–225.7% [63]	MMSE [mean (SD)]: n.r. ICC = 0.57–0.93 MDC$_{95\%}$ = n.r [63].	MMSE [mean (SD)]: 13.8 (5.7) ICC = 0.96–0.99 MDC$_{95\%}$ = 19.6–33.3% [43]
BBS		MMSE [mean (SD)]: 20.0 (5.5) ICC = 0.95 MDC$_{95\%}$ = 38.6% [14]		MMSE [mean (SD)]: 13.8 (5.7) ICC = 0.99 MDC$_{95\%}$ = 10.2% [43]
6 m WT	MMSE [mean (SD)]: 22.7 (2.1) ICC = 0.83 MDC$_{95\%}$ = 41.5% [17]	MMSE [mean (SD)]: 19.2 (4.4) ICC = 0.86 MDC$_{95\%}$ = 36.5% [17]	MMSE [mean (SD)]: 15.5 (2.4) ICC = 0.89 MDC$_{95\%}$ = 31.6% [17]	MMSE [mean (SD)]: 16.9 (7.3) ICC = 0.80–0.95 MDC$_{95\%}$ = n.r [102].
5x STS	MMSE [mean (SD)]: 21.4 (5.0) ICC = 0.80 MDC$_{95\%}$ = 29.9% [51]			MMSE [mean (SD)]: 16.9 (7.3) ICC = 0.94 MDC$_{95\%}$ = n.r [102].
Modified 30s CST	MMSE [mean (SD)]: 22.7 (2.1) ICC = 0.79 MDC$_{95\%}$ = 45.7% [17]	MMSE [mean (SD)]: 19.2 (4.4) ICC = 0.84 MDC$_{95\%}$ = 42.5% [17]	MMSE [mean (SD)]: 15.5 (2.4) ICC = 0.88 MDC$_{95\%}$ = 33.2% [17]	

5x STS Five Times Sit-to-Stand Test, *6 m WT* 6-m walk test, *30s CST* 30-s chair stand test, *BBS* Berg Balance Scale, *FICSIT-4* Frailty and Injuries: Cooperative Studies of Intervention Techniques - subtest 4, *GMWT* Groningen Meander Walking Test, *ICC* Intraclass correlation coefficient, *MDC$_{95\%}$* Percentage minimal detectable changes at 95% confidence interval, *MMSE* Mini-Mental State Examination, *n.r.* Not reported, *SD* Standard deviation.

Frequency of use and effect sizes of motor assessments applied in previous randomised controlled trials

TUG, BBS, 5x STS, POMA, 30s CST, and instrumented gait analysis, were the most frequently applied assessments, utilised in six to 16 RCT. We were only able to calculate effect sizes for 12 studies, as F/t statistics and/ or standard deviations of baseline-post differences were infrequently reported. Effect sizes were large for FR, BBS, POMA, TUG, instrumented gait analysis, 5x STS, ACSID, and 30s CST (see Table 1/Additional file 9 for motor assessments identified during first search without available information on psychometric properties).

Summary and derivation of recommendations

Aiming to derive comprehensive recommendations on motor assessments for IWD, we combined the results of primary and secondary outcomes for each physical domain as summarised in Table 7.

Considering all information on primary and secondary outcomes, the derived recommendations include the following motor assessments:

– Balance: FR, GMWT (time), BBS, and POMA
– Mobility and gait: TUG and instrumented gait analysis to assess spatiotemporal gait parameters

Table 5 Subgroup analysis of test-retest reliability considering aetiology of dementia

	Alzheimer's disease	Various types/not reported
BBS	ICC = 0.95 MDC$_{95\%}$ = 38.6% [14]	ICC = 0.99 MDC$_{95\%}$ = 10.2% [43]
TUG (between-day reliability)	ICC = 0.72–0.76 (MDC$_{95\%}$ = 20.3–24.9%) [14, 51]	ICC = 0.87–0.99 (MDC$_{95\%}$ = 15.8–39.6%) [17, 43, 102]
Up and go tasks (between-day reliability)	ICC = 0.51–0.76 (MDC$_{95\%}$ = 20.3–36.2%) [14, 51]	ICC = 0.87–0.99 (MDC$_{95\%}$ = 15.8–39.6%) [17, 43, 102]
Short distance walking speed (without NeuroCom Balance Master)	ICC = 0.95–0.98 MDC$_{95\%}$ = 10.2–28.9% [6, 121]	ICC = 0.83–0.95 MDC$_{95\%}$ = 31.6–84.3% [17, 26, 43]
5x STS	ICC = 0.80 MDC$_{95\%}$ = 29.9% [51]	ICC = 0.94 MDC$_{95\%}$ = n.r [102].
STS assessments (without Rising Index)	ICC = 0.02–0.80 MDC$_{95\%}$ = 29.9–80.2% [51]	ICC = 0.79–0.94 MDC$_{95\%}$ = 33.2–45.7% [17, 102]
Maximum isometric strength assessed with dynamometers	ICC = 0.95–0.98 MDC$_{95\%}$ = n.r [114].	ICC = 0.63–0.71 MDC$_{95\%}$ = n.r [102].

5x STS Five Times Sit-to-Stand Test, *BBS* Berg Balance Scale, *ICC* Intraclass correlation coefficient, *MDC$_{95\%}$* Percentage minimal detectable changes at 95% confidence interval, *n.r.* Not reported, *STS* Sit-to-Stand, *TUG* Timed Up & Go Test.

Table 6 Subgroup analysis of test-retest reliability considering cueing

	No cueing	Verbal cueing or verbal and visual/tactile cueing	More extensive cueing including physical assistance
FR		ICC = 0.84 MDC$_{95\%}$ = 15.4% [51]	ICC = 0.81 MDC$_{95\%}$ = 68.9% [14]
GMWT	ICC = 0.57–0.96 MDC$_{95\%}$ = 31.2–225.7% [63]		ICC = 0.96–0.99 MDC$_{95\%}$ = 19.6–33.3% [43]
TUG		ICC = 0.76–0.96 MDC$_{95\%}$ = 23.3–39.6% [17, 51, 102]	ICC = 0.72–0.99 MDC$_{95\%}$ = 15.8–30.0% [6, 14, 43]
Short distance WT		ICC = 0.80–0.95 MDC$_{95\%}$ = 31.6–41.5% [17, 102]	ICC = 0.85 MDC$_{95\%}$ = 84.3% [43]
Short distance walking speed	ICC = 0.95–0.96 MDC$_{95\%}$ = 10.2–12.0% [121]	ICC = 0.50–0.95 MDC$_{95\%}$ = 31.6–48.3% [17, 26, 51]	ICC = 0.85–0.98 MDC$_{95\%}$ = 25.5–84.3% [6, 43]

FR Functional Reach Test, *GMWT* Groningen Meander Walking Test, *ICC* Intraclass correlation coefficient, *MDC$_{95\%}$* Percentage minimal detectable changes at 95% confidence interval, *TUG* Timed Up & Go Test, *WT* Walk tests.

– Strength: STS assessments with more than one repetition
– Endurance: 6 min WT
– Functional Performance: No recommendation possible, due to insufficient research on psychometric properties

These recommendations are based on several outcomes rated in the highest category or one outcome rated in the highest and at least two in the second category (see Table 7).

Discussion

We addressed the purpose of this systematic review to quantitatively examine motor assessments for IWD by comprehensively analysing psychometric properties (primary outcome), frequency of use, and effect sizes (secondary outcomes) in a two-stage literature search. Recommendations on motor assessments are based on primary and secondary outcomes. Additionally, we analysed the influence of severity and aetiology of dementia and cueing on test-retest reliability.

Findings on primary and secondary outcomes

The systematic search identified only few investigations examining validity, internal consistency, and intra-rater reliability of motor assessments in IWD. Thus, we were not able to draw further conclusions or consider these outcomes for deriving recommendations. Summarizing findings for inter-rater reliability shows sufficient relative inter-rater reliability and relatively low MDC$_{95\%}$ of considered motor assessments. Hence, they are objective measures to determine motor performance in IWD. Motor assessments analysing time in tasks of short duration, such as 4 m WT, should, however, be treated with caution, as small measurement errors may significantly influence absolute inter-rater reliability. With respect to

test-retest reliability, the majority of identified investigations observed sufficient relative test-retest reliability, while absolute test-retest reliability was mainly indeterminate or unacceptable. This supports their usage to investigate changes on a group level, but does not allow assessing intra-individual changes [7, 17, 31]. Moreover, decreasing test-retest reliability from between-day, through within-day, to within-session investigations may be related to fluctuating daily forms in IWD. We expect that characteristics of daily form, such as mood or motivational aspects, remain relatively constant within short intervals, while they potentially alter with increasing time. More research is necessary to develop criteria to determine daily form, aiming to ensure comparable conditions in longitudinal investigations. Besides, fluctuating daily forms in IWD may have contributed to observed unacceptable absolute test-retest reliability. Other explanations refer to high intra-individual variability in IWD and related inappropriate or naive selection of metrics, which do not account for this variability.

Regarding frequency of use, previous trials predominately applied clinical motor assessments established in healthy older adults or various clinical populations, while those considering specific characteristics of IWD such as GMWT, Physiomat, or ACSID, were less frequently applied. This may be related to their first introduction between 2014 and 2018. Due to insufficient information in previous RCT, we were only able to determine time*-group interaction effect sizes for 38% of analysed motor assessments. Based on large effect sizes reported in at least one RCT, we assumed sensitivity to change for most of these assessments.

Findings on influence of severity and aetiology of dementia and cueing on test-retest reliability

Considering severity of dementia, we expected decreasing test-retest reliability with increasing cognitive impairment.

Table 7 Summary of outcomes to derive recommendations for motor assessments for individuals with dementia

Motor assessment	Inter-rater reliability		Test-retest reliability		Frequency of use	Time*group interaction effect size
	relative	absolute	relative	absolute		
Balance						
FICSIT-4	?	?	0	–	–	?
Modified Clinical Test of Sensory Interaction of Balance	?	?	0	–	–	?
Limits of Stability	?	?	–	–	–	?
Physiomat-Trail-Making Task	?	?	0	?	–	?
Physiomat-Follow-The-Ball Task	?	?	0	?	–	?
FR	0	?	0	0	0	+
Hill Step Test	?	?	0	0	0	?
Step Quick Turn Test	?	?	–	–	–	?
Figure of Eight Test	?	?	0	–	–	?
GMWT	0	0	+	0	–	?
BBS	+	0	+	0	+	+
Modified BBS	?	?	?	?	–	?
POMA	?	?	?	?	+	+
Mobility and gait						
TUG	+	+	+	+	+	+
Cognitive TUG	?	?	–	0	0	?
Manual TUG	?	?	+	0	–	?
6 m WT	0	0	0	–	0	?
4 m WT	0	–	0	–	0	0
Instrumented gait analysis	?	?	0	0	+	+
Strength						
5x STS	?	?	0	0	+	+
STS on NeuroCom Balance Master	?	?	–	–	–	?
ACSID	+	?	?	?	–	+
30s CST	0	0	0	–	+	+
Handgrip dynamometer	?	?	+	–	0	–
Maximum isometric strength assessed with dynamometers	?	?	0	?	0	?
Endurance						
6 min WT	0	?	+	0	0	?
Functional performance						
SPPB	?	?	?	?	0	0
E-ADL Test	?	?	0	?	0	?

4 m WT 4-m walk test, *5x STS* Five Times Sit-to-Stand Test, *6 m WT* 6-m walk test, *6 min WT* 6-min walk test, *30s CST* 30-s chair stand test, *ACSID* Assessment of Compensatory Sit-to-Stand Maneuvers in People With Dementia, *BBS* Berg Balance Scale, *E-ADL Test* Erlangen Test of Activities of Daily Living, *FICSIT-4* Frailty and Injuries: Cooperative Studies of Intervention Techniques - subtest 4, *FR* Functional Reach Test, *GMWT* Groningen Meander Walking Test, *POMA* Performance Oriented Mobility Assessment, *SPPB* Short Physical Performance Battery, *STS* Sit-to-Stand, *TUG* Timed Up & Go Test.
Relative reliability: - = insufficient, 0 = sufficient, very low/low quality of evidence, + = sufficient, moderate/high quality of evidence,? = not investigated.
Absolute reliability: - = insufficient/unacceptable, 0 = indeterminate/inconsistent/sufficient, very low/low quality of evidence, + = sufficient, moderate/high quality of evidence,? = not investigated.
Frequency of use: - = 1 randomised controlled trial, 0 = 2–5 randomised controlled trials, += > 5 randomised controlled trials.
*Time*group interaction effect size*: - = no effect, 0 = at least one trial with small or medium effect, + = at least one trial with large effect,? = could not be calculated/not reported.

This assumption was true for FICSIT-4 and GMWT but not for all assessments. Severity of dementia may only influence specific assessments, for example those with complex instructions or assessing outcomes frequently impaired in IWD, such as balance [10]. Unexpectedly, we observed increasing test-retest reliability with increasing

severity of dementia for BBS, 6 m WT, modified 30s CST, and 5x STS. However, these observations were only based on single studies, which partly differed in characteristics, such as aetiology of dementia.

Regarding the aetiology of dementia, test-retest reliability of BBS and up and go tasks was lower for AD than for various or not reported types. Both assessments consist of several short tasks and include multi-step instructions. Compared to other aetiologies, individuals with AD may have more difficulties in understanding and/or remembering such instructions, which potentially influences test-retest reliability [14, 23, 122]. In contrast, test-retest reliability of walking speed was higher in AD which could be related to later occurring gait impairments in AD [20]. Additional research on aetiologies, however, is required to understand lower test-retest reliability of STS tasks and higher test-retest reliability of maximum isometric strength assessed with dynamometers in AD.

Analysing the influence of cueing on test-retest reliability revealed higher test-retest reliability when cueing was allowed or more extensive for GMWT and TUG, which are assessments consisting of unfamiliar or several short tasks. Cueing possibly stabilises motor performance by supporting impaired cognitive performance and thus improves test-retest reliability. In contrast, short distance WT, for which test-retest reliability was higher when cueing was not allowed or less extensive, are close to everyday life, include single-stage tasks, and consider well automated movement processes not requiring additional cognitive support. Accordingly, cueing rather may distract IWD leading to destabilised performance decreasing test-retest reliability. No explanation for the same association in FR is available.

Based on these observed influences, we derived the following suggestions:

- Put emphasis on simple instructions, especially for IWD with advanced stages or AD.
- Consider individual cognitive and motor deficits, when selecting motor assessments.
- Only use cueing for motor assessments where it is inevitable.

Recommendations and need for future research
Recommendations for balance assessments include FR, GMWT (time), BBS, and POMA. Due to infrequent use and insufficient research on psychometric properties, feasibility and sensitivity to change of GMWT and psychometric properties of POMA require further investigation. Focusing on mobility and gait, we suggest to apply TUG and spatiotemporal gait parameters assessed with instrumented gait analysis. Comparing different gait analysis systems, NeuroCom Balance Master, however, seems to

be less suitable. Despite insufficient or equivocal results, future research should investigate short distance WT of different distances, as instrumented gait analysis systems may not be available for all studies. Considering strength, we suggest to apply STS assessments comprising more than one repetition, which, however, predominantly determine functional performance of lower limbs. Thus, further evaluation of strength assessments including upper limb strength and measures allowing conclusion on actual strength performance are required. Moreover, we suggest to use the 6 min WT as an endurance assessment for IWD. Future research on endurance assessment, however, is crucial since this was the only identified assessment. As information on psychometric properties is insufficient, we are not able to recommend any functional performance assessment. Based on secondary outcomes some indications are available for SPPB. However, psychometric properties of SPPB and other functional performance assessments need to be investigated in future studies.

Comparison with state of research
Recommendations of motor assessments in this review are largely in line with those of previous reviews [13, 24]. Small discrepancies may be related to distinctions in identified assessments and studies, different prioritisation of considered outcomes, and divergent criteria for good measurement properties. Additionally, this review, consistently to Fox et al. [7], determined sufficient relative test-retest reliability for the majority of motor assessments in IWD, but remarked high $MDC_{95\%}$ reflecting unacceptable absolute test-retest reliability.

Similarly, motor assessments recommended in this review are mainly in line with those elaborated in a qualitative approach [22]. However, FICSIT, 6 m WT, SPPB, and Physical Performance Test were rated appropriate in the qualitative approach, but could not be recommended based on quantitative outcomes as they were infrequently used or insufficiently investigated. Further discrepancies on FR, which was rated inappropriate but can be recommended based on quantitative outcomes, require additional examination. Moreover, some general indications, related to consideration of specific characteristics and cueing are consistently suggested. Accordingly, this review largely sustains the recommendations elaborated in a qualitative approach.

General considerations on primary and secondary outcomes
The interpretation of findings regarding psychometric properties is challenging as there are no firm criteria for acceptable reliability in literature [31]. Regardless of concrete criteria, ICC do not only reflect relative reliability but also can be related to sample size or variability in the sample [123]. Accordingly, trial-to-trial consistency

can be poor, despite high ICC. Thus, it is advised not to focus on single estimates of reliability and to additionally consider absolute reliability [17, 31]. Due to lack of information on minimal important change of motor assessments in IWD, we could scarcely apply COSMIN criteria for absolute reliability. Besides, Smidt et al. [42] arbitrarily defined that a difference of 10% in minimal detectable change would be acceptable. Other research groups referred to them and introduced another cut-off of 30% without any justification [43, 44]. In absence of other criteria, we adopted this cut-off of 30% to identify unacceptable $MDC_{95\%}$ but not to conclude on sufficient absolute reliability.

Frequency of use and effect sizes do not necessarily allow conclusions to be drawn on quality of motor assessments and should not be overestimated. Regardless of appropriateness and meaningfulness, researchers may decide to apply motor assessments as they are commonly used or easy to utilise. Nonetheless, frequency of use can provide indications about feasibility of motor assessments, which is based on the assumption that unfeasible motor assessments do not disseminate as good as feasible ones. Comparably, effect sizes can provide information on sensitivity to change, but are also dependent on effectiveness of interventions.

Strengths and limitations
To our knowledge, this is the first systematic review utilising a comprehensive approach combining different outcomes of previous reviews by performing an extensive two-stage literature search. We need to state potential risk of bias regarding the selection of considered motor assessments. Due to restricting the analysis of motor assessments to those applied in RCT, some assessments may be missing. Furthermore, large heterogeneity of included psychometric property studies limits the meaningfulness of derived recommendations. As psychometric properties are potentially influenced by various determinants, such as sample size, sample characteristics including severity and aetiology of dementia, cueing, test-retest interval, or considered outcomes, we cannot ensure that the deductions on psychometric properties are true and not randomly caused by differing determinants. Therefore, false assumptions, undetected influences or relations, and random observations may have occurred. Similarly, the consideration of several influences on test-retest reliability only allows rough estimations, which could be also affected by heterogeneity of analysed studies. Moreover, insufficient information on execution of motor assessments, severity and aetiology of dementia, and cueing in available investigations impeded detailed analyses and limited meaningfulness of observations. Accordingly, the elaborated recommendations should be used with care and further research investigating psychometric properties and dementia specific influences on test-retest reliability is required.

Conclusion
Despite the necessity for further research in various areas, this review establishes an important foundation for future investigations. Additionally, direct implications for studies determining effectiveness of physical activity on motor performance in IWD can be derived. However, elaborated recommendations cannot be considered as final conclusions since the analysis of primary and secondary outcomes reveals several challenges and areas of insufficient research, and only focus on quantitative aspects. Furthermore, new assessments, especially developed for IWD, are required. Such assessments can be based on prior tasks but should consider specific characteristics of IWD. Additionally, it is of high importance to standardise motor assessments and cueing to ensure comparability between studies. Herein, standardisation refers to selection and performance procedures of motor assessments and external cues. Currently, a wide range of motor assessments (e.g. previous RCT applied 19 different balance assessments) with different performance procedures (e.g. different ratings or modifications) as well as various external cues (e.g. clearly defined verbal cues vs. as much assistance as needed) are frequently applied to determine the same motor functions or quantities. Accordingly, recommendations on specific motor assessments as well as indications on assessment procedures elaborated in quantitative and qualitative (see [22]) approaches are important to improve standardisation. Evidence on effectiveness of physical activity can contribute to gain access to physical activity interventions and thereby positively influence quality of life in IWD. Determining evidence, however, is not possible without appropriate, sensitive, valid, reliable, and standardised motor assessments, which consider the individual characteristics of single individuals.

Supplementary information

> **Additional file 1.** Search term first search.
>
> **Additional file 2.** Formulas for calculating time*group interaction effect sizes.
>
> **Additional file 3.** Search term main search.
>
> **Additional file 4.** Formulas for calculating minimal detectable change at 95% confidence interval.
>
> **Additional file 5.** Study characteristics first search.
>
> **Additional file 6.** Study characteristics main search.
>
> **Additional file 7.** Data extraction first search.
>
> **Additional file 8.** Data extraction main search.
>
> **Additional file 9.** Description, frequency of use, and effect sizes of motor assessments applied in previous randomised controlled trials without available information on psychometric properties.

Abbreviations
30 s CST: 30-s chair stand test; 4 m WT: 4-m walk test; 5x STS: Five Times Sit-to-Stand Test; 6 m WT: 6-m walk test; 6 min WT: 6-min walk test;

ACSID: Assessment of Compensatory Sit-to-Stand Maneuvers in People With Dementia; AD: Alzheimer's disease; BBS: Berg Balance Scale; COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments; E-ADL Test: Erlangen Test of Activities of Daily Living; FICSIT-4: Frailty and Injuries: Cooperative Studies of Intervention Techniques - subtest 4; FR: Functional Reach Test; GMWT: Groningen Meander Walking Test; ICC: Intraclass correlation coefficient/s; IWD: Individuals with dementia; $MDC_{95\%}$: Percentage minimal detectable change/s at 95% confidence interval; MMSE: Mini-Mental State Examination; POMA: Performance Oriented Mobility Assessment; RCT: Randomised controlled trial/s; SPPB: Short Physical Performance Battery; STS: Sit-to-Stand/sit-to-stand; TUG: Timed Up & Go Test; WT: Walk test/s

Author details
[1]Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Engler-Bunte-Ring 15, 76131 Karlsruhe, Germany. [2]Faculty of Kinesiology, University of Calgary, Calgary, Alberta, Canada.

References
1. Ahlskog JE, Geda YE, Graff-Radford NR, Petersen RC. Physical exercise as a preventive or disease-modifying treatment of dementia and brain aging. Mayo Clin Proc. 2011;86:876–84. https://doi.org/10.4065/mcp.2011.0252.
2. Hauer K, Becker C, Lindemann U, Beyer N. Effectiveness of physical training on motor performance and fall prevention in cognitively impaired older persons: a systematic review. Am J Phys Med Rehabil. 2006;85:847–57. https://doi.org/10.1097/01.phm.0000228539.99682.32.
3. Brett L, Traynor V, Stapley P. Effects of physical exercise on health and well-being of individuals living with a dementia in nursing homes: a systematic review. J Am Med Dir Assoc. 2016;17:104–16. https://doi.org/10.1016/j.jamda.2015.08.016.
4. Gonçalves A-C, Cruz J, Marques A, Demain S, Samuel D. Evaluating physical activity in dementia: a systematic review of outcomes to inform the development of a core outcome set. Age Ageing. 2018;47:34–41. https://doi.org/10.1093/ageing/afx135.
5. Wittwer JE, Webster KE, Hill K. Reproducibility of gait variability measures in people with Alzheimer's disease. Gait Posture. 2013;38:507–10. https://doi.org/10.1016/j.gaitpost.2013.01.021.
6. Ries JD, Echternach JL, Nof L, Gagnon BM. Test-retest reliability and minimal detectable change scores for the timed "up & go" test, the six-minute walk test, and gait speed in people with Alzheimer disease. Phys Ther. 2009;89:569–79. https://doi.org/10.2522/ptj.20080258.
7. Fox B, Henwood T, Keogh J, Neville C. Psychometric viability of measures of functional performance commonly used for people with dementia: a systematic review of measurement properties. JBI Database System Rev Implement Rep. 2016;14:115–71. https://doi.org/10.11124/JBISRIR-2016-003064.
8. Baddeley A, Logie R, Bressi S, Della Sala S, Spinnler H. Dementia and working memory. Quart J Exper Psychol Sect A. 1986;38:603–18. https://doi.org/10.1080/14640748608401616.
9. Perry RJ, Hodges JR. Attention and executive deficits in Alzheimer's disease. A critical review. Brain. 1999;122(Pt 3):383–404. https://doi.org/10.1093/brain/122.3.383.
10. Allan LM, Ballard CG, Burn DJ, Kenny RA. Prevalence and severity of gait disorders in Alzheimer's and non-Alzheimer's dementias. J Am Geriatr Soc. 2005;53:1681–7. https://doi.org/10.1111/j.1532-5415.2005.53552.x.
11. Manckoundia P, Mourey F, Pfitzenmeyer P, Papaxanthis C. Comparison of motor strategies in sit-to-stand and back-to-sit motions between healthy and Alzheimer's disease elderly subjects. Neuroscience. 2006;137:385–92. https://doi.org/10.1016/j.neuroscience.2005.08.079.
12. van Iersel MB, Hoefsloot W, Munneke M, Bloem BR, Olde Rikkert MGM. Systematic review of quantitative clinical gait analysis in patients with dementia. Z Gerontol Geriatr. 2004;37:27–32. https://doi.org/10.1007/s00391-004-0176-7.
13. Bossers WJR, van der Woude LHV, Boersma F, Scherder EJA, van Heuvelen MJG. Recommended measures for the assessment of cognitive and physical performance in older patients with dementia: a systematic review. Dement Geriatr Cogn Dis Extra. 2012;2:589–609. https://doi.org/10.1159/000345038.
14. Muir-Hunter SW, Graham L, Montero OM. Reliability of the Berg balance scale as a clinical measure of balance in community-dwelling older adults with mild to moderate Alzheimer disease: a pilot study. Physiother Can. 2015;67:255–62. https://doi.org/10.3138/ptc.2014-32.
15. Telenius EW, Engedal K, Bergland A. Inter-rater reliability of the Berg balance scale, 30 s chair stand test and 6 m walking test, and construct validity of the Berg balance scale in nursing home residents with mild-to-moderate dementia. BMJ Open. 2015;5:e008321. https://doi.org/10.1136/bmjopen-2015-008321.
16. Hauer K, Oster P. Measuring functional performance in persons with dementia. J Am Geriatr Soc. 2008;56:949–50. https://doi.org/10.1111/j.1532-5415.2008.01649.x.
17. Blankevoort CG, van Heuvelen MJG, Scherder EJA. Reliability of six physical performance tests in older people with dementia. Phys Ther. 2013;93:69–78. https://doi.org/10.2522/ptj.20110164.
18. Phillips CD, Chu CW, Morris JN, Hawes C. Effects of cognitive impairment on the reliability of geriatric assessments in nursing homes. J Am Geriatr Soc. 1993;41:136–42. https://doi.org/10.1111/j.1532-5415.1993.tb02047.x.
19. Cohen-Mansfield J. Heterogeneity in dementia: challenges and opportunities. Alzheimer Dis Assoc Disord. 2000;14:60–3.
20. Valkanova V, Ebmeier KP. What can gait tell us about dementia? Review of epidemiological and neuropsychological evidence. Gait Posture. 2017;53:215–23. https://doi.org/10.1016/j.gaitpost.2017.01.024.
21. van Iersel MB, Benraad CEM, Olde Rikkert MGM. Validity and reliability of quantitative gait analysis in geriatric patients with and without dementia. J Am Geriatr Soc. 2007;55:632–4. https://doi.org/10.1111/j.1532-5415.2007.01130.x.
22. Trautwein S, Barisch-Fritz B, Scharpf A, Bossers W, Meinzer M, Steib S, et al. Recommendations for assessing motor performance in individuals with dementia: suggestions of an expert panel – a qualitative approach. Eur Rev Aging Phys Act. 2019. https://doi.org/10.1186/s11556-019-0212-7.
23. Lee H-S, Park S-W. The reliability of balance, gait, and muscle strength test for the elderly with dementia: a systematic review. KSPM. 2017;12:49–58. https://doi.org/10.13066/kspm.2017.12.3.49.
24. McGough EL, Lin S-Y, Belza B, Becofsky KM, Jones DL, Liu M, et al. A scoping review of physical performance outcome measures used in exercise interventions for older adults with Alzheimer disease and related dementias. J Geriatr Phys Ther. 2019;42:28–47. https://doi.org/10.1519/JPT.0000000000000159.
25. Lundin-Olsson L, Nyberg L, Gustafson Y. Attention, frailty, and falls: the effect of a manual task on basic mobility. J Am Geriatr Soc. 1998;46:758–61. https://doi.org/10.1111/j.1532-5415.1998.tb03813.x.
26. McGough EL, Logsdon RG, Kelly VE, Teri L. Functional mobility limitations and falls in assisted living residents with dementia: physical performance

assessment and quantitative gait analysis. J Geriatr Phys Ther. 2013;36:78–86. https://doi.org/10.1519/JPT.0b013e318268de7f.

27. Beauchet O, Allali G, Berrut G, Hommet C, Dubost V, Assal F. Gait analysis in demented subjects: interests and perspectives. Neuropsychiatr Dis Treat. 2008;4:155–60.

28. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6:e1000097. https://doi.org/10.1371/journal.pmed.1000097.

29. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6:e1000100. https://doi.org/10.1371/journal.pmed.1000100.

30. 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;63:737–45. https://doi.org/10.1016/j.jclinepi.2010.02.006.

31. Bruton A, Conway JH, Holgate ST. Reliability: what is it, and how is it measured? Physiotherapy. 2000;86:94–9. https://doi.org/10.1016/S0031-9406(05)61211-4.

32. Carter R, Lubinsky J, Domholdt E. Rehabilitation research: principles and applications. 4th ed. St. Louis, Missouri: Elsevier Health Sciences; 2013.

33. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep. 1985;100:126–31.

34. Thalheimer W, Cook S. How to calculate effect sizes from published research: a simplified methodology; 2002.

35. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: L. Erlbaum Associates; 1988.

36. 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;12:189–98. https://doi.org/10.1016/0022-3956(75)90026-6.

37. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27:1171–9. https://doi.org/10.1007/s11136-017-1765-4.

38. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27:1147–57. https://doi.org/10.1007/s11136-018-1798-3.

39. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Upper Saddle River: Pearson/Prentice Hall; 2015.

40. Schwenk M, Gogulla S, Englert S, Czempik A, Hauer K. Test-retest reliability and minimal detectable change of repeated sit-to-stand analysis using one body fixed sensor in geriatric patients. Physiol Meas. 2012;33:1931–46. https://doi.org/10.1088/0967-3334/33/11/1931.

41. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs): user manual; 2018.

42. Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Devillé WL, de Winter AF, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. Arch Phys Med Rehabil. 2002;83:1145–50. https://doi.org/10.1053/apmr.2002.33728.

43. Lee HS, Park SW, Chung HK. The Korean version of relative and absolute reliability of gait and balance assessment tools for patients with dementia in day care center and nursing home. J Phys Ther Sci. 2017;29:1934–9. https://doi.org/10.1589/jpts.29.1934.

44. Huang S-L, Hsieh C-L, Wu R-M, Tai C-H, Lin C-H, Lu W-S. Minimal detectable change of the timed "up & go" test and the dynamic gait index in people with Parkinson disease. Phys Ther. 2011;91:114–21. https://doi.org/10.2522/ptj.20090126.

45. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October. 2013;2013.

46. Forbes D, Forbes SC, Blake CM, Thiessen EJ, Forbes S. Exercise programs for people with dementia. Cochrane Database Syst Rev. 2015:CD006489. https://doi.org/10.1002/14651858.CD006489.pub4.

47. Feldman HH, Woodward M. The staging and assessment of moderate to severe Alzheimer disease. Neurology. 2005;65:S10–7. https://doi.org/10.1212/WNL.65.6_suppl_3.S10.

48. Hogan DB, Bailey P, Carswell A, Clarke B, Cohen C, Forbes D, et al. Management of mild to moderate Alzheimer's disease and dementia. Alzheimers Dement. 2007;3:355–84. https://doi.org/10.1016/j.jalz.2007.07.006.

49. Rossiter-Fornoff JE, Wolf SL, Wolfson LI, Buchner DM. A cross-sectional validation study of the FICSIT common Data Base static balance measures: cooperative studies of intervention techniques. J Gerontol Ser A-Biol Sci Med Sci. 1995;50A:M291–7. https://doi.org/10.1093/gerona/50A.6.M291.

50. Bossers WJR, van der Woude LHV, Boersma F, Hortobágyi T, Scherder EJA, van Heuvelen MJG. A 9-week aerobic and strength training program improves cognitive and motor function in patients with dementia: a randomized, controlled trial. Am J Geriatr Psychiatry. 2015;23:1106–16. https://doi.org/10.1016/j.jagp.2014.12.191.

51. Suttanon P, Hill KD, Dodd KJ, Said CM. Retest reliability of balance and mobility measurements in people with mild to moderate Alzheimer's disease. Int Psychogeriatr. 2011;23:1152–9. https://doi.org/10.1017/S1041610211000639.

52. Suttanon P, Hill KD, Said CM, Williams SB, Byrne KN, LoGiudice D, et al. Feasibility, safety and preliminary evidence of the effectiveness of a home-based exercise programme for older people with Alzheimer's disease: a pilot randomized controlled trial. Clin Rehabil. 2013;27:427–38. https://doi.org/10.1177/0269215512460877.

53. Wiloth S, Lemke N, Werner C, Hauer K. Validation of a computerized, game-based assessment strategy to measure training effects on motor-cognitive functions in people with dementia. JMIR Ser Games. 2016;4:e12. https://doi.org/10.2196/games.5696.

54. Wiloth S, Werner C, Lemke NC, Bauer J, Hauer K. Motor-cognitive effects of a computerized game-based training method in people with dementia: a randomized controlled trial. Aging Ment Health. 2018;22:1124–35. https://doi.org/10.1080/13607863.2017.1348472.

55. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. J Gerontol. 1990;45:M192–7. https://doi.org/10.1093/geronj/45.6.M192.

56. Arcoverde C, Deslandes A, Moraes H, Almeida C, NBd A, Vasques PE, et al. Treadmill training as an augmentation treatment for Alzheimer's disease: a pilot randomized controlled study. Arq Neuropsiquiatr. 2014;72:190–6. https://doi.org/10.1590/0004-282X20130231.

57. Miu D, Szeto S, Mak Y. A randomized controlled trial on the effect of exercise on physical, cognitive, and affective function in dementia subjects. Asian J Gerontol Geriatr. 2008;3:8–16.

58. Netz Y, Axelrad S, Argov E. Group physical activity for demented older adults feasibility and effectiveness. Clin Rehabil. 2007;21:977–86. https://doi.org/10.1177/0269215507078318.

59. Vreugdenhil A, Cannell J, Davies A, Razay G. A community-based exercise programme to improve functional ability in people with Alzheimer's disease: a randomized controlled trial. Scand J Caring Sci. 2012;26:12–9. https://doi.org/10.1111/j.1471-6712.2011.00895.x.

60. Hill KD. A new test of dynamic standing balance for stroke patients: reliability, validity and comparison with healthy elderly. Physiother Can. 1996;48:257–62. https://doi.org/10.3138/ptc.48.4.257.

61. Wesson J, Clemson L, Brodaty H, Lord S, Taylor M, Gitlin L, et al. A feasibility study and pilot randomised trial of a tailored prevention program to reduce falls in older people with mild dementia. BMC Geriatr. 2013;13:89. https://doi.org/10.1186/1471-2318-13-89.

62. Johansson G, Jarnlo G-B. Balance training in 70-year-old women. Physiother Theory Pract. 2009;7:121–5. https://doi.org/10.3109/09593989109106962.

63. Bossers WJR, van der Woude LHV, Boersma F, Scherder EJA, van Heuvelen MJG. The Groningen meander walking test: a dynamic walking test for older adults with dementia. Phys Ther. 2014;94:262–72. https://doi.org/10.2522/ptj.20130077.

64. Berg K. Measuring balance in the elderly: preliminary development of an instrument. Physiother Can. 1989;41:304–11. https://doi.org/10.3138/ptc.41.6.304.

65. Burgener SC, Yang Y, Gilbert R, Marsh-Yant S. The effects of a multimodal intervention on outcomes of persons with early-stage dementia. Am J Alzheimers Dis Other Demen. 2008;23:382–94. https://doi.org/10.1177/1533317508317527.

66. Christofoletti G, Oliani MM, Gobbi S, Stella F, Bucken Gobbi LT, Renato CP. A controlled clinical trial on the effects of motor intervention on balance and cognition in institutionalized elderly patients with dementia. Clin Rehabil. 2008;22:618–26. https://doi.org/10.1177/0269215507086239.

67. Kim M-J, Han C-W, Min K-Y, Cho C-Y, Lee C-W, Ogawa Y, et al. Physical exercise with multicomponent cognitive intervention for older adults with Alzheimer's disease: a 6-month randomized controlled trial. Dement Geriatr Cogn Dis Extra. 2016;6:222–32. https://doi.org/10.1159/000446508.
68. Lam FMH, Liao LR, Kwok TCY, Pang MYC. Effects of adding whole-body vibration to routine day activity program on physical functioning in elderly with mild or moderate dementia: a randomized controlled trial. Int J Geriatr Psychiatry. 2018;33:21–30. https://doi.org/10.1002/gps.4662.
69. Padala KP, Padala PR, Lensing SY, Dennis RA, Bopp MM, Roberson PK, et al. Home-based exercise program improves balance and fear of falling in community-dwelling older adults with mild Alzheimer's disease: a pilot study. J Alzheimers Dis. 2017;59:565–74. https://doi.org/10.3233/JAD-170120.
70. Padala KP, Padala PR, Malloy TR, Geske JA, Dubbert PM, Dennis RA, et al. Wii-fit for improving gait and balance in an assisted living facility: a pilot study. J Aging Res. 2012;2012:597573. https://doi.org/10.1155/2012/597573.
71. Telenius EW, Engedal K, Bergland A. Effect of a high-intensity exercise program on physical function and mental health in nursing home residents with dementia: an assessor blinded randomized controlled trial. PLoS One. 2015;10:e0126102. https://doi.org/10.1371/journal.pone.0126102.
72. Toots A, Littbrand H, Lindelöf N, Wiklund R, Holmberg H, Nordström P, et al. Effects of a high-intensity functional exercise program on dependence in activities of daily living and balance in older adults with dementia. J Am Geriatr Soc. 2016;64:55–64. https://doi.org/10.1111/jgs.13880.
73. Yoon JE, Lee SM, Lim HS, Kim TH, Jeon JK, Mun MH. The effects of cognitive activity combined with active extremity exercise on balance, walking activity, memory level and quality of life of an older adult sample with dementia. J Phys Ther Sci. 2013;25:1601–4. https://doi.org/10.1589/jpts.25.1601.
74. Dawson N, Judge KS, Gerhart H. Improved functional performance in individuals with dementia after a moderate-intensity home-based exercise program: a randomized controlled trial. J Geriatr Phys Ther. 2019;42:18–27. https://doi.org/10.1519/JPT.0000000000000128.
75. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. J Am Geriatr Soc. 1986;34:119–26. https://doi.org/10.1111/j.1532-5415.1986.tb05480.x.
76. Francese T, Sorrell J, Butler FR. The effects of regular exercise on muscle strength and functional abilities of late stage Alzheimer's residents. Am J Alzheimers Dis Other Demen. 1997;12:122–7. https://doi.org/10.1177/153331759701200305.
77. Hauer K, Ullrich P, Dutzi I, Beurskens R, Kern S, Bauer J, et al. Effects of standardized home training in patients with cognitive impairment following geriatric rehabilitation: a randomized controlled pilot study. Gerontology. 2017;63:495–506. https://doi.org/10.1159/000478263.
78. Hauer K, Schwenk M, Zieschang T, Essig M, Becker C, Oster P. Physical training improves motor performance in people with dementia: a randomized controlled trial. J Am Geriatr Soc. 2012;60:8–15. https://doi.org/10.1111/j.1532-5415.2011.03778.x.
79. Kovács E, Sztruhár Jónásné I, Karóczi CK, Korpos A, Gondos T. Effects of a multimodal exercise program on balance, functional mobility and fall risk in older adults with cognitive impairment: a randomized controlled single-blind study. Eur J Phys Rehabil Med. 2013;49:639–48.
80. Santana-Sosa E, Barriopedro MI, López-Mojares LM, Pérez M, Lucia A. Exercise training is beneficial for Alzheimer's patients. Int J Sports Med. 2008;29:845–50. https://doi.org/10.1055/s-2008-1038432.
81. Podsiadlo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39:142–8. https://doi.org/10.1111/j.1532-5415.1991.tb01616.x.
82. Cancela JM, Ayán C, Varela S, Seijo M. Effects of a long-term aerobic exercise intervention on institutionalized patients with dementia. J Sci Med Sport. 2016;19:293–8. https://doi.org/10.1016/j.jsams.2015.05.007.
83. Kampragkou C, Iakovidis P, Kampragkou E, Kellis E. Effects of a 12-week aerobic exercise program combined with music therapy and memory exercises on cognitive and functional ability in people with middle type of Alzheimer's disease. Int J Physiother. 2017. https://doi.org/10.15621/ijphy/2017/v4i5/159420.
84. Sobol NA, Hoffmann K, Frederiksen KS, Vogel A, Vestergaard K, Brændgaard H, et al. Effect of aerobic exercise on physical performance in patients with Alzheimer's disease. Alzheimers Dement. 2016;12:1207–15. https://doi.org/10.1016/j.jalz.2016.05.004.
85. Toulotte C, Fabre C, Dangremont B, Lensel G, Thévenon A. Effects of physical training on the physical capacity of frail, demented patients with a history of falling: a randomised controlled trial. Age Ageing. 2003;32:67–73. https://doi.org/10.1093/ageing/32.1.67.
86. Aguiar P, Monteiro L, Feres A, Gomes I, Melo A. Rivastigmine transdermal patch and physical exercises for Alzheimer's disease: a randomized clinical trial. Curr Alzheimer Res. 2014;11:532–7. https://doi.org/10.2174/1567205011666140618102224.
87. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the timed up & go test. Phys Ther. 2000;80:896–903. https://doi.org/10.1093/ptj/80.9.896.
88. Guralnik JM, Seeman TE, Tinetti ME, Nevitt MC, Berkman LF. Validation and use of performance measures of functioning in a non-disabled older population: MacArthur studies of successful aging. Aging (Milano). 1994;6:410–9. https://doi.org/10.1007/BF03324272.
89. Rolland Y, Pillard F, Klapouszczak A, Reynish E, Thomas D, Andrieu S, et al. Exercise program for nursing home residents with Alzheimer's disease: a 1-year randomized, controlled trial. J Am Geriatr Soc. 2007;55:158–65. https://doi.org/10.1111/j.1532-5415.2007.01035.x.
90. de Souto Barreto P, Cesari M, Denormandie P, Armaingaud D, Vellas B, Rolland Y. Exercise or social intervention for nursing home residents with dementia: a pilot randomized, controlled trial. J Am Geriatr Soc. 2017;65:E123–9. https://doi.org/10.1111/jgs.14947.
91. Toots A, Littbrand H, Holmberg H, Nordström P, Lundin-Olsson L, Gustafson Y, et al. Walking aids moderate exercise effects on gait speed in people with dementia: a randomized controlled trial. J Am Med Dir Assoc. 2017;18:227–33. https://doi.org/10.1016/j.jamda.2016.09.003.
92. Kressig RW, Beauchet O. Guidelines for clinical applications of spatio-temporal gait analysis in older adults. Aging Clin Exp Res. 2006;18:174–6. https://doi.org/10.1007/BF03327437.
93. Pedrinolla A, Venturelli M, Fonte C, Munari D, Benetti MV, Rudi D, et al. Exercise training on locomotion in patients with Alzheimer's disease: a feasibility study. J Alzheimers Dis. 2018;61:1599–609. https://doi.org/10.3233/JAD-170625.
94. Schwenk M, Dutzi I, Englert S, Micol W, Najafi B, Mohler J, et al. An intensive exercise program improves motor performances in patients with dementia: translational model of geriatric rehabilitation. J Alzheimers Dis. 2014;39:487–98. https://doi.org/10.3233/JAD-130470.
95. Schwenk M, Zieschang T, Englert S, Grewal G, Najafi B, Hauer K. Improvements in gait characteristics after intensive resistance and functional training in people with dementia: a randomised controlled trial. BMC Geriatr. 2014;14:73. https://doi.org/10.1186/1471-2318-14-73.
96. Kemoun G, Thibaud M, Roumagne N, Carette P, Albinet C, Toussaint L, et al. Effects of a physical training programme on cognitive function and walking efficiency in elderly persons with dementia. Dement Geriatr Cogn Disord. 2010;29:109–14. https://doi.org/10.1159/000272435.
97. Csuka M, McCarty DJ. Simple method for measurement of lower extremity muscle strength. Am J Med. 1985;78:77–81. https://doi.org/10.1016/0002-9343(85)90465-6.
98. Steinberg M, Leoutsakos J-MS, Podewils LJ, Lyketsos CG. Evaluation of a home-based exercise program in the treatment of Alzheimer's disease: the maximizing Independence in dementia (MIND) study. Int J Geriatr Psychiatry. 2009;24:680–5. https://doi.org/10.1002/gps.2175.
99. Werner C, Wiloth S, Lemke NC, Kronbach F, Hauer K. Development and validation of a novel motor-cognitive assessment strategy of compensatory sit-to-stand maneuvers in people with dementia. J Geriatr Phys Ther. 2018;41:143–54. https://doi.org/10.1519/JPT.0000000000000116.
100. Werner C, Wiloth S, Lemke NC, Kronbach F, Jansen C-P, Oster P, et al. People with dementia can learn compensatory movement maneuvers for the sit-to-stand task: a randomized controlled trial. J Alzheimers Dis. 2017;60:107–20. https://doi.org/10.3233/JAD-170258.
101. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport. 1999;70:113–9. https://doi.org/10.1080/02701367.1999.10608028.
102. Thomas VS, Hageman PA. A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia. Int Psychogeriatr. 2002;14:17–23. https://doi.org/10.1017/S1041610202008244.
103. Verkerke GJ, Lemmink KAPM, Slagers AJ, Westhoff MH, van Riet GAJ, Rakhorst G. Precision, comfort and mechanical performance of the Quadriso-tester, a quadriceps force measuring device. Med Biol Eng Comput. 2003;41:283–9. https://doi.org/10.1007/BF02348432.
104. Enright PL. The six-minute walk test. Respir Care. 2003;48:783–5.
105. Roach KE, Tappen RM, Kirk-Sanchez N, Williams CL, Loewenstein D. A randomized controlled trial of an activity specific exercise program for individuals with Alzheimer disease in long-term care settings. J Geriatr Phys Ther. 2011;34:50–6. https://doi.org/10.1519/JPT.0b013e31820aab9c.

106. Tappen RM, Roach KE, Applegate EB, Stowell P. Effect of a combined walking and conversation intervention on functional mobility of nursing home residents with Alzheimer disease. Alzheimer Dis Assoc Disord. 2000; 14:196–201. https://doi.org/10.1097/00002093-200010000-00002.

107. Venturelli M, Scarsini R, Schena F. Six-month walking program changes cognitive and ADL performance in patients with Alzheimer. Am J Alzheimers Dis Other Demen. 2011;26:381–8. https://doi.org/10.1177/1533317511418956.

108. 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:M85–94. https://doi.org/10.1093/geronj/49.2.M85.

109. Pitkälä KH, Pöysti MM, Laakkonen M-L, Tilvis RS, Savikko N, Kautiainen H, et al. Effects of the Finnish Alzheimer disease exercise trial (FINALEX): a randomized controlled trial. JAMA Intern Med. 2013;173:894–901. https://doi.org/10.1001/jamainternmed.2013.359.

110. Graessel E, Viegas R, Stemmer R, Küchly B, Kornhuber J, Donath C. The Erlangen test of activities of daily living: first results on reliability and validity of a short performance test to measure fundamental activities of daily living in dementia patients. Int Psychogeriatr. 2009;21:103–12. https://doi.org/10.1017/S1041610208007710.

111. Luttenberger K, Schmiedeberg A, Gräßel E. Activities of daily living in dementia: revalidation of the E-ADL test and suggestions for further development. BMC Psychiatry. 2012;12:208. https://doi.org/10.1186/1471-244X-12-208.

112. Bossers WJR, van der Woude LHV, Boersma F, Hortobágyi T, Scherder EJA, van Heuvelen MJG. Comparison of effect of two exercise programs on activities of daily living in individuals with dementia: a 9-week randomized, controlled trial. J Am Geriatr Soc. 2016;64:1258–66. https://doi.org/10.1111/jgs.14160.

113. Henskens M, Nauta IM, Drost KT, Scherder EJ. The effects of movement stimulation on activities of daily living performance and quality of life in nursing home residents with dementia: a randomized controlled trial. Clin Interv Aging. 2018;13:805–17. https://doi.org/10.2147/CIA.S160031.

114. Suzuki M, Yamada S, Inamura A, Omori Y, Kirimoto H, Sugimura S, et al. Reliability and validity of measurements of knee extension strength obtained from nursing home residents with dementia. Am J Phys Med Rehabil. 2009;88:924–33. https://doi.org/10.1097/PHM.0b013e3181ae1003.

115. Bronas UG, Salisbury D, Kelly K, Leon A, Chow L, Yu F. Determination of aerobic capacity via cycle ergometer exercise testing in Alzheimer's disease. Am J Alzheimers Dis Other Demen. 2017;32:500–8. https://doi.org/10.1177/1533317517720065.

116. Suzuki M, Kirimoto H, Inamura A, Yagi M, Omori Y, Yamada S. The relationship between knee extension strength and lower extremity functions in nursing home residents with dementia. Disabil Rehabil. 2012;34:202–9. https://doi.org/10.3109/09638288.2011.593678.

117. Schwenk M, Hauer K, Zieschang T, Englert S, Mohler J, Najafi B. Sensor-derived physical activity parameters can predict future falls in people with dementia. Gerontology. 2014;60:483–92. https://doi.org/10.1159/000363136.

118. Tappen RM, Roach KE, Buchner D, Barry C, Edelstein J. Reliability of physical performance measures in nursing home residents with Alzheimer's disease. J Gerontol Ser A-Biol Sci Med Sci. 1997;52A:M52–5. https://doi.org/10.1093/gerona/52A.1.M52.

119. van Iersel MB, Munneke M, Esselink RAJ, Benraad CEM, Olde Rikkert MGM. Gait velocity and the timed-up-and-go test were sensitive to changes in mobility in frail elderly patients. J Clin Epidemiol. 2008;61:186–91. https://doi.org/10.1016/j.jclinepi.2007.04.016.

120. Alencar MA, Dias JMD, Figueiredo LC, Dias RC. Handgrip strength in elderly with dementia: study of reliability. Rev Bras Fisioter. 2012;16:510–4. https://doi.org/10.1590/S1413-35552012005000059.

121. Wittwer JE, Webster KE, Andrews PT, Menz HB. Test-retest reliability of spatial and temporal gait parameters of people with Alzheimer's disease. Gait Posture. 2008;28:392–6. https://doi.org/10.1016/j.gaitpost.2008.01.007.

122. Orange JB, Molloy DW, Lever JA, Darzins P, Ganesan CR. Alzheimer's disease. Physician-patient communication. Can Fam Physician. 1994;40:1160–8.

123. Koo TK, Li MY. A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

Publisher's Note