JOURNAL OF SPORTS SCIENCES
2025, VOL. 43, NO. 16, 1616–1628
https://doi.org/10.1080/02640414.2025.2516178

Routledge
Taylor & Francis Group

PHYSICAL ACTIVITY, HEALTH AND EXERCISE

Check for updates

Establishing a core outcome set to evaluate the effects of power training in older adults: A Delphi study

Mohamed El Hadouchi[a,b,c,d], Jaap van Dieën[b], Francois Maissan[a], Timon Eskes[c], Cindy Veenhof[d,e] and Henri Kiers[a,b]

[a]Institute for Human Movement Sciences, University of Applied Sciences Utrecht, Utrecht, The Netherlands; [b]Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [c]Health Research Consultancy, Sports & Science Health, Lochem, The Netherlands; [d]Research Group Innovation of Human Movement Care, University of Applied Sciences Utrecht, Utrecht, The Netherlands; [e]Department of Rehabilitation, Physical Therapy Science and Sport, Brain Center, University Medical Center Utrecht, Utrecht, The Netherlands

ABSTRACT

Studies evaluating the effects of power training in older adults use a variety of measurement tools and outcome parameters, limiting comparability of results and calling the validity of conclusions into question. This study aimed to identify a core outcome set (COS) to measure the effects of power training in older adults, encompassing the function, activities, and participation domain of the International Classification of Functioning, Disease, and Health (ICF). Twenty-one tests were evaluated based on their ability to measure muscle power and the effects of power training. Our methodology consisted of two Delphi survey rounds and an expert panel meeting using modified Nominal Group Technique. The COS consisted of tests considered most feasible for clinical practice and least burdensome for older adults. The COS included the Squat Jump Test (paired with accelerometery-based instrumentation) and the Timed Up-and-Go Test for the function and function and activities domain, respectively. No test was identified for the participation domain, however, experts proposed using daily-life accelerometery until more suitable outcome measures are defined and validated. This study addresses a critical research gap in standardised assessment protocols, and contributes to a multifaceted approach to measuring the impact of power training in older adults.

ARTICLE HISTORY
Received 6 January 2025
Accepted 29 May 2025

KEYWORDS
Power training; older adults; core outcome set; Delphi; consensus

Introduction

As the global population ages, it becomes increasingly important to implement effective strategies to sustain physical functioning and foster functional independence among older adults (Ferrucci et al., 2016; World Health Organization, 2017; World Health Organization, 2017). One such strategy focuses on a crucial element of physical functioning, namely, muscle power, which reflects the capacity of the muscles to generate high levels of force while contracting rapidly (Bassey et al., 1992; Bean et al., 2009; Tieland et al., 2018). Previous research has named muscle power as a discriminative variable in the natural decline of physical functioning that is associated with ageing and pathological conditions, such as sarcopenia (Fielding et al., 2011). Muscle power declines faster with age than muscle strength (Delmonico et al., 2009; Metter et al., 1997; Skelton et al., 1995; Strotmeyer et al., 2018; Wiegmann et al., 2021), and is essential for maintaining and improving physical functioning as it directly influences an individual's ability to perform daily activities and engage in physical exercise (Bean et al., 2002; Fielding et al., 2011). Examples from everyday life where muscle power plays an important role include activities such as rising from a chair, climbing a flight of stairs, or lifting objects from the ground. Muscle power also allows for quick reactions to prevent falls after stumbling or losing balance (Cadore et al., 2013), and may thus have implications for balance and fall risk during the completion of activities of daily living.

One of the most effective strategies for maintaining and improving muscle power is progressive power training, a type of exercise focused on quick muscle contractions and explosive movements (Foldvari et al., 2000; Miszko et al., 2003). Studies show that older adults who participate in power training, particularly when targeting the lower extremities, demonstrate enhanced muscular power as well as improved physical functioning

CONTACT Mohamed El Hadouchi Mohamed.elhadouchi@hu.nl, mohamed@powerful-ageing.nl Institute for Human Movement Sciences, University of Applied Sciences Utrecht, P.O. Box 12011, Utrecht 3501 AA, The Netherlands

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02640414.2025.2516178

© 2025 Informa UK Limited, trading as Taylor & Francis Group

and physical performance (Balachandran et al., 2014; el; Hadouchi et al., 2022; Steib et al., 2010; Tschopp et al., 2011). However, the interpretation of these findings is hindered by several limitations. First, studies evaluating the effects of power training in older adults used a wide selection of measurement tools and outcome parameters, which limits direct comparison of results between studies. A recent review found consistent challenges in the reliability, validity, and responsiveness of several tests commonly used to assess the effects of power training in older adults, including the Functional Fait Assessment and the Four Step Square Test (el Hadouchi et al., 2024). This highlights the need for further research to identify measurement instruments with sufficient clinimetric properties. Second, there is a discrepancy between the domains in which outcomes are measured and conclusions are drawn. For instance, muscle power may be measured at the body function level of the International Classification of Functioning, Disease, and Health (ICF), but interpretations are often made at the activity or participation level, without clear justification for this extrapolation. Changes in muscle power may influence an individual's ability to perform specific tasks (activities) and engage in daily life and societal activities (participation) (World Health Organization, 2001), but this depends on many factors beyond muscle power alone and requires careful evaluation.

The interconnected nature of ICF health domains necessitates a comprehensive approach to assessing the effects of power training on physical functioning in older adults. In this study, we aimed to reach consensus on a core outcome set (COS) of measurement instruments that can be used to evaluate the effects of power training in older adults. This COS should cover the function, activities, and participation domains of the ICF.

Methods

Study design

The development of a COS to measure the effects of power training in older adults was achieved using two consensus group methodologies, namely the Delphi technique for survey rounds 1 and 2, followed by an expert panel meeting using a modified Nominal Group Technique (m-NGT) (Humphrey-Murto et al., 2017; Jones & Hunter, 1995). The NGT is a method for structuring expert meetings aimed at gathering information and achieving consensus on a pre-specified topic, with a specific focus on a singular objective (Jones & Hunter, 1995; McMillan et al., 2016). In this study, adjustments were made to the manner in which experts were assigned tasks to facilitate a better evaluation and categorisation of measurement instruments across the ICF domains. This is explained more thoroughly under the header titled expert panel meeting.

The study utilised a workgroup comprising five members, including two members from the research team (MeH and FM) and three clinical experts specialising in geriatric care (TE, EN, and RP). Ethical approval and participant consent for our Delphi process and expert meeting was not required based on the Dutch Medical Research Involving Human Subjects Act (WMO).

Selection of participants

Participants were recruited within the professional networks of Vrije Universiteit Amsterdam and Hogeschool Utrecht (University of Applied Sciences Utrecht). Additional participants were enlisted by consulting established experts in the field, who recommended peers with relevant expertise. This snowball sampling method allowed for a wide spectrum of expertise across all stakeholders involved in training and treatment of older adults, ensuring a diverse and well-balanced representation of knowledge and skills. To be eligible for inclusion, participants had to have: a) a minimum of 5 years of clinical experience in physical therapy or experience conducting research related to power training, exercise physiology, sports performance, physiotherapy care, or geriatrics; and b) possession of a postgraduate degree (Master or PhD) in one of the aforementioned fields or a related discipline, and specialised in one of the areas under investigation (Hasson et al., 2000; Keeney et al., 2006).

The recruitment process consisted of three phases: (1) distribution of formal written invitations to identified experts; (2) a two-week deliberation period; and (3) personalised follow-up via email and telephone to those who had not responded. We invited 20 participants for Delphi survey rounds 1 and 2 and anticipated to receive between 12 to 16 responses (Humphrey-Murto et al., 2017).

Study procedure

Upon accepting the invitation, study participants received an introduction letter and instructions for completing the digital Delphi survey between November 2022 and March 2023. Participants were further provided with information regarding the objective and construct of each test, the theoretical framework related to the concepts and ICF domains, and relevant terminology (Appendix 1). Reminder

calls and emails were sent after 4 and 6 weeks if the survey had not yet been returned. The survey remained open for a total of 12 weeks to maximise the response rate, but any surveys received after this period were excluded from the analysis. This study's procedural framework is summarised in Figure 1.

Delphi survey rounds 1 and 2

The Delphi survey consisted of two rounds. In the first round, the starting point was formed by the 21 tests identified in a previous systematic review (el Hadouchi et al., 2024). Participants were asked to: (a) respond to the suitability of the test for measuring the effects of

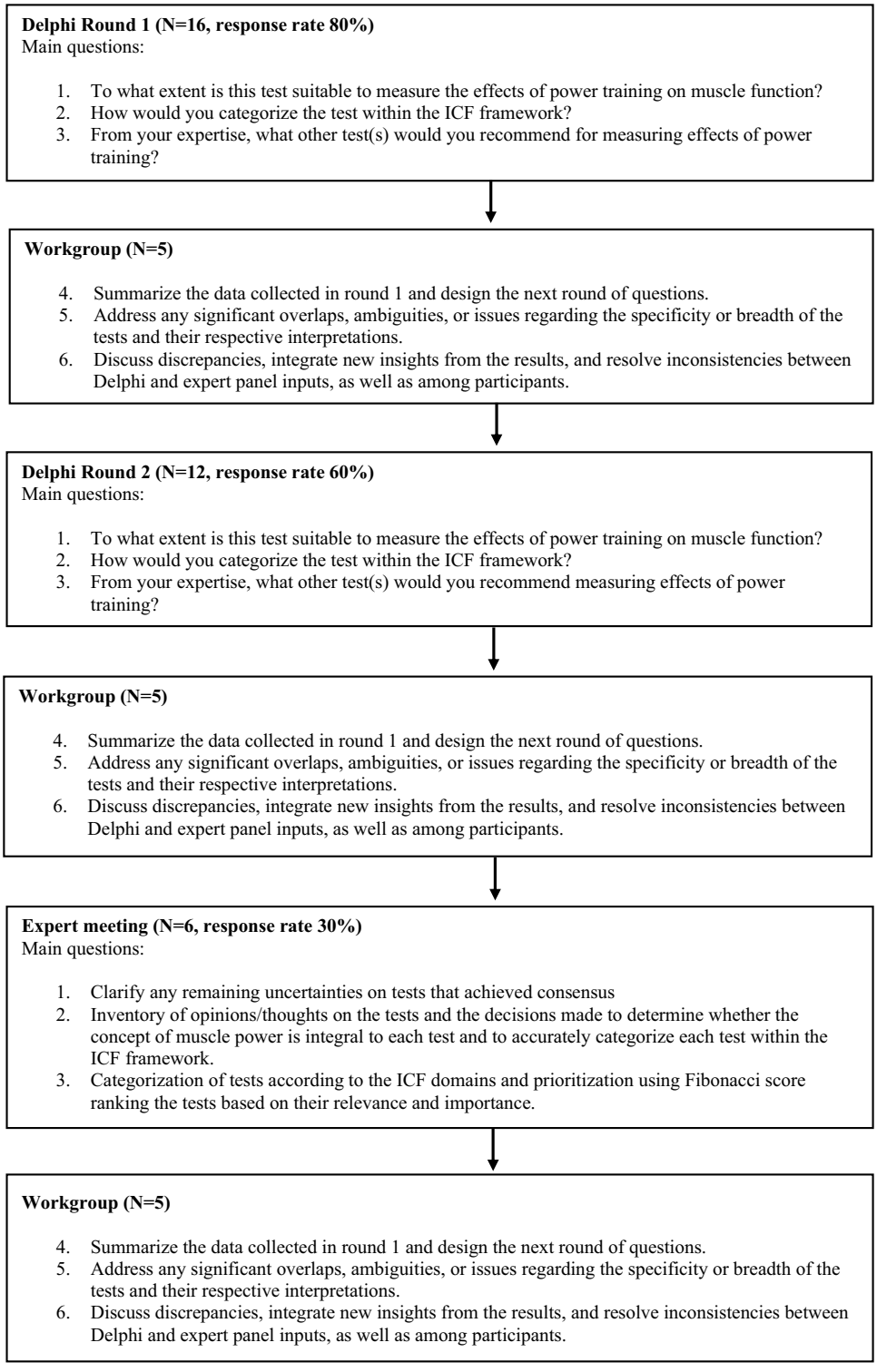

Figure 1. Flowchart of study procedural framework. The figure 1 that was uploaded with the original manuscript had the arrows in the middle of each of the boxes. Could these arrows be moved to the middle again so that the figure is symmetric? Thank you

power training; (b) identify which health domain of the ICF the test assessed; and (c) share any additional tests that they recommended for measuring (the effects of) power training. Following the first Delphi round, the workgroup summarised the data: tests with low levels of agreement (<60%) were excluded from further analyses; tests with uncertain levels of agreement (between 60–70%) or tests that were suggested by study participants in round 1 were carried forward to round 2 of the Delphi survey; and tests with > 70% agreement were considered to have reached consensus (Al Juffali et al., 2019; Mokkink et al., 2010; Vergouw et al., 2011; Vogel et al., 2019), and were carried forward to the expert panel meeting.

The second round of the Delphi survey was structured and executed in the same manner as the first round, but more detailed explanations were provided for each test to resolve any unclarities that could have contributed to the uncertain levels of agreement. Additionally, participant were asked to which extent the tests were suitable for measuring the concept of muscle power. Following Delphi round 2, the workgroup summarised the data: tests with low levels of agreement (<60%) were excluded from further analysis, while the remainder of tests (both tests with uncertain levels of agreement (60–70%) and tests that achieved consensus (<70%) were carried forward to the expert panel meeting. In both rounds of the Delphi survey, participants had the opportunity to submit remarks to explain their choices or share any ambiguities. These remarks were used by the workgroup to improve the work process, and also influenced the selection of tests carried forward to the next round (i.e., if several participants shared their doubt regarding a certain test, this test was carried forward to the next round regardless of its level of agreement).

Expert panel meeting

All Delphi survey participants were invited to the expert panel meeting organised in June 2023. Additionally, three members of the research team (MeH, FM and TE) joined the meeting to support the process, but were explicitly instructed not to sway the participants (Humphrey-Murto et al., 2017). Prior to the expert panel meeting, participants received a digital file containing a summary of the Delphi survey results, as this knowledge was the starting point for the meeting.

Following a brief introduction and explanation of the purpose of the expert panel meeting, the structure of the meeting was as follows: (a) discussion on whether each tests could measure the effect of power training, and if so, categorisation of the test within the ICF domain(s); (b) arrangement of the tests into quadrants defined by two axes (ICF quadrant). In Figure 2, the horizontal axis represents the health domain of the ICF, while the vertical axis differentiates the activities domain in terms of capacity and performance; and (3) prioritisation of tests within ICF quadrants using a Fibonacci Scoring principle (McMillan et al., 2016). To identify the most important tests for measuring the effects of power training, each participant independently assigned Fibonacci scores (choosing from the following values: 1, 2, 3, 5, 8, 13, 21, 34 and 55) to a test based on: (1) the test's ability to measure the concept muscle power; (2) the test's feasibility for use in clinical practice; and (3) the test's feasibility in older adults. If tests received equal scores, the workgroup selected the test that was least burdensome.

Results

Study participants

Of the 20 invited experts, 16 (80% response rate) responded to the first round of the Delphi survey and 12 (75% response rate compared to round 1) responded to the second round of the Delphi survey. These participants included academics ($n = 4$), sports physiotherapists ($n = 4$), geriatric physiotherapists ($n = 4$), exercise scientists ($n = 2$), and clinical exercise physiologists ($n = 2$). The four participants who dropped out of the study following the first survey round did so due to limited availability. The names of all 16 panel members are mentioned in the acknowledgements section, and participant characteristics are summarised in Table 1.

Delphi survey round 1

Of the 21 tests included in the first round of the Delphi survey, nine tests reached consensus (agreement level > 70%), three tests had uncertain agreement scores (60–-70%), and the remaining nine tests had low agreement scores (<60%) regarding their suitability for measuring the effects of power training (Table 2). The nine tests that reached consensus were classified within the activities domain of the ICF. A complete overview of agreement scores per test can be found in Appendix 2 and Appendix 3.

A total of eight tests were carried over to the second round of the Delphi survey, which consisted of the three tests with uncertain agreement scores (the Stair Climb Test, 30-sec Functional Strength Test, and 400-meter Walking Test) and an additional four tests submitted by participants in Delphi survey round 1 (Burpee Test, GymAware Linear Positional Transducer, Margaria-Kalamen Power Test, and daily-life accelerometry). Additionally, some participants expressed doubts

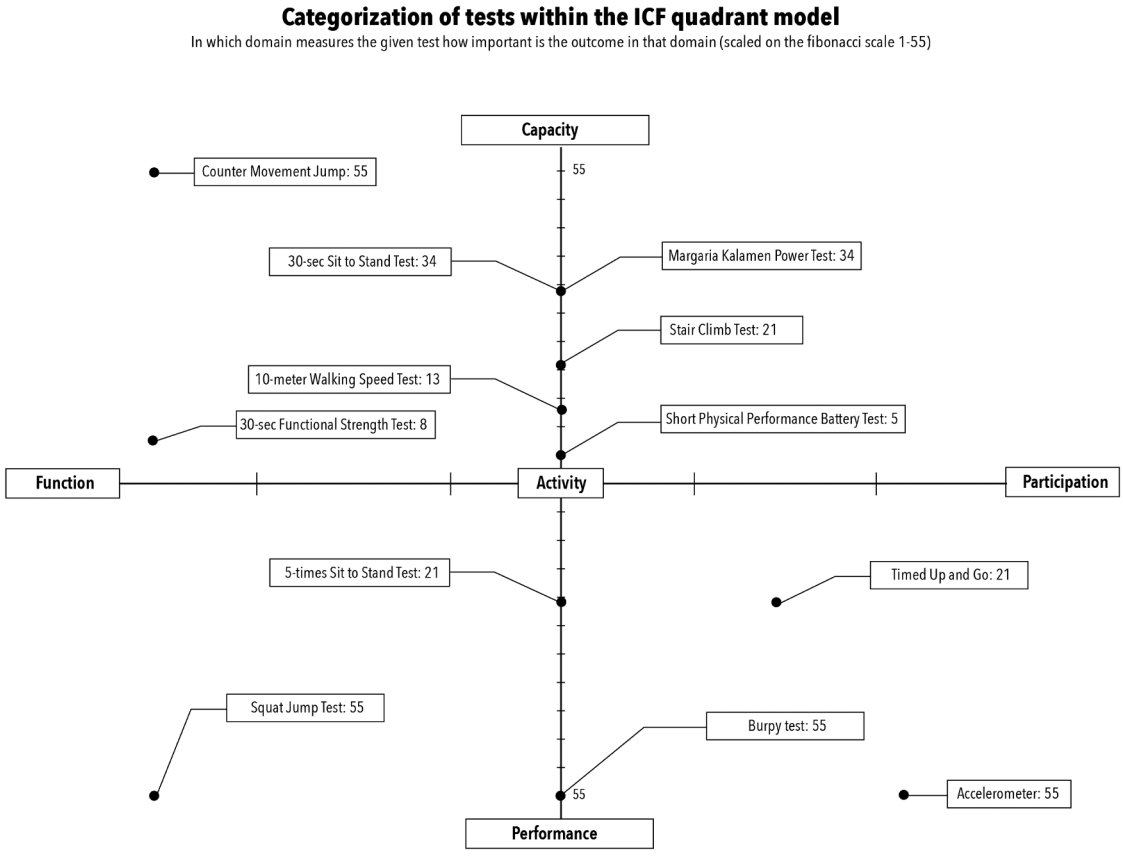

Figure 2. Allocation of the tests within the ICF quadrant model.

Table 1. Participant characteristics ($N = 16$).

Characteristics	N (%)
Gender, male	8 (50%)
Education	
PhD or PhD candidate	8 (50%)
MSc	8 (50%)
Healthcare providers (sport and geriatric physiotherapist)	7 (44%)
Combination of research, lecture and clinical practice	9 (56%)
Occupation at the time of the study	
Academic	4 (25%)
Sports physiotherapist	4 (25%)
Geriatric physiotherapist	4 (25%)
Exercise scientist	2 (13%)
Clinical exercise physiologist	2 (13%)

regarding the use of the Sit to Rise test. Despite a clear description of the test in the survey, different interpretations had led to questions surrounding the test's ability to measure muscle power. To clear any ambiguities among respondents, the Sit to Rise test was also included in the second round of the Delphi survey.

Delphi survey round 2

Of the eight tests included in round 2 of the Delphi survey, five tests reached consensus (agreement level > 70%) and three tests had low agreement (<60%) regarding their suitability for measuring the effects of power training (Table 3). One of the tests with low agreement, accelerometry, received several remarks by participants. The ability of the accelerometer to measure a wide range of physical activities and movements patterns led to confusion regarding its consistent use and application. To clarify its use among participants, this test was carried forward to the expert panel meeting. A complete overview of agreement scores per test can be found in Appendix 4 and 5.

After two Delphi survey rounds, a total of 15 tests was identified and carried over to the expert panel meeting, of which 14 tests achieved consensus regarding their ability to measure the effects of power training (Table 4).

Table 2. Delphi survey round 1.

#	Test	Agreement on the suitability of the test for measuring the effects of power training	Extent of agreement across ICF domains	Consensus achieved?
1	5-times Sit to Stand Test	100%	Function: 6% Activity: 94% Participation : 0%	YES
2	30-sec Sit to Stand Test	94%	Function: 6% Activity: 94% Participation: 0%	YES
3	Timed Up and Go Test/8Foot Up and Go Test	94%	Function: 0% Activity: 94% Participation: 6%	YES
4	Squat Jump Test	81%	Function: 25% Activity: 75% Participation: 0%	YES
5	Counter Movement Jump	81%	Function: 31% Activity: 69% Participation: 0%	YES
6	10-meter Walking Speed Test	75%	Function: 0% Activity: 100% Participation: 0%	YES
7	Medicine Ball Thow Test	70%	Function: 12% Activity: 88% Participation: 0%	YES
8	Short Physical Performance Battery Test	70%	Function: 0% Activity: 100% Participation: 0%	YES
9	Wingate Test	70%	Function: 25% Activity: 75% Participation: 0%	YES
10	Stair Climb Test	69%	Function: 6% Activity: 82% Participation: 12%	UNCERTAIN
11	30-sec Functional Strength Test	63%	Function: 63% Activity: 37% Participation: 0%	UNCERTAIN
12	400-meter Walking Test	60%	Function: 0% Activity: 88% Participation: 12%	UNCERTAIN
13	The Four Square Step Test	59%	Function: 0% Activity: 94% Participation: 6%	NO
14	1RM Test	56%	Function: 88% Activity: 12% Participation: 0%	NO
15	Isokinetic Dynamometer	56%	Function : 100% Activity: 0% Participation: 0%	NO
16	Continuous Scale Physical Functional Performance Test	50%	Function: 0% Activity: 88% Participation: 12%	NO
17	Sit to Rise Test	50%	Function: 0% Activity: 88% Participation: 12%	NO, BUT NEEDS DISCUSSION
18	6-min Walk Test	38%	Function: 0% Activity: 88% Participation: 12%	NO
19	6- &10-meter Timed Walk Test	38%	Function: 6% Activity: 88% Participation: 6%	NO
20	Functional Gait Assessment	38%	Function: 0% Activity: 88% Participation: 12%	NO
21	Timed Balance Test	31%	Function: 12% Activity: 81% Participation: 6%	NO

Table 3. Delphi survey round 2.

		Suitability of the test for measuring the effects of power training			Suitability of the test for measuring the concept muscle power		
#	Test	Agreement	Extent of agreement across ICF domains	Consensus achieved?	Agreement	Extent of agreement on "speed" across ICF domains	Consensus achieved?
1	GymAware Linear Position Transducer	100%	Function: 50% Activity: 50% Participation: 0%	YES	100%	Power: 33% Activity emphasizing speed: 42%	YES
2	Margaria Kalamen Power Test	100%	Function: 8% Activity: 92% Participation: 0%	YES	92%	Power: 8% Activity emphasizing speed: 58%	YES
3	30-sec Functional Strength Test	91%	Function: 75% Activity: 25% Participation: 0%	YES	73%	Power: 58% Activity emphasizing speed: 8%	YES
4	Burpee Test	75%	Function: 0% Activity: 100% Participation: 0%	YES	67%	Power 0% Activity emphasizing speed: 75%	UNCERTAIN
5	Stair Climb Test	75%	Function: 0% Activity: 100% Participation: 0%	YES	58%	Power: 8% Activity emphasizing speed: 50%	UNCERTAIN
6	Sit to Rise Test	58%	Function: 8% Activity: 84% Participation: 8%	NO	17%	Power: 8% Activity emphasizing speed: 8%	NO
7	400-meter Walk Test	42%	Function: 8% Activity: 92% Participation: 0%	NO	8%	Power: 8% Activity emphasizing speed: 42%	NO
8	Accelerometry	33%	Function: 16% Activity: 42% Participation: 42%	NO, BUT NEEDS DISCUSSION	33%	Power: 16% Activity emphasizing speed: 0%	NO

Table 4. Tests identified following two Delphi survey rounds.

#	Test
1	Short Physical Performance Battery Test
2	30-sec Sit to Stand Test
3	5-times Sit to Stand Test
4	Counter Movement Jump
5	Medicine Ball Thow Test
6	Timed Up and Go/8-Foot Up and Go Test
7	Squat Jump Test
8	Wingate Test
9	10-meter Walking Speed Test
10	Stair Climb Test
11	30-sec Functional Strength Test
12	GymAware Linear Position Transducer
13	Burpee Test
14	Margaria Kalamen Power Test
15	Daily-life accelerometery

Expert panel meeting

Of the 16 invited study participants, six attended the expert panel meeting (response rate: 37.5%). Attendees were sports physiotherapists ($n = 2$), geriatric physiotherapists ($n = 2$), academic ($n = 1$), and a clinical exercise physiologists ($n = 1$). In addition, three workgroup members (MeH, FM and TE) joined to support the process.

During group deliberations, the panel of experts reached a consensus on 11 tests regarding their ability to measure the effects of power training in older adults. The remaining four tests (accelerometry, GymAware Linear Power Transducer, Medicine Ball Throw, and the Wingate test) were excluded for various reasons. The accelerometer and the GymAware Linear Power Transducer were considered measurement tools intended to measure physical activity, and unlike performance evaluations that directly gauge physical capabilities through structured challenges, they collect data on movement patterns. They thus provide insight into the intensity and duration of activities rather than being standalone tests. The Medicine Ball Throw test was not ideally suited for assessing the effects of power training in older adults as this test requires a certain level of coordination and motor skills, which may not align with age-related declines in motor control and coordination. Also, the Medicine Ball Throw Test does not reflect

daily physical activities commonly engaged in by most older adults. The Wingate test, while effective for measuring anaerobic capacity, was considered to be less suitable for assessing muscle power in older adults as it requires maximal effort, is demanding, and can be potentially dangerous for older individuals with age-related physical limitations. The test's focus on brief, high-intensity performance does not reflect the sustained, moderate-intensity muscle power needed for performing activities of daily life, making it less appropriate for older adults.

The 11 tests for which consensus was reached were categorised into the ICF quadrant model (Figure 2). The Counter Movement Jump and Squat Jump were categorised under the function domain, both receiving a score of 55 on the Fibonacci scale. The remaining tests were categorised within the activity or participation domain, and further classified into sub-domains based on whether they focused on generic capacity or capacity with an emphasis on speed. The highest scoring tests were the Timed Up-and Go-Test and the Margaria Kalamen Power Test with 34 points each, and the Stair Climb Test, and 30-seconds Sit-to-Stand with 21 points each, and; the 5-times Sit-to-Stand and 10-meter Walking Speed Test scored 13 points each. Participants agreed that the Counter Movement Jump, the Squat Jump Test, 10-meter Walking Speed Test, and the Stair Climb Test were specific for measuring muscle power.

Following group deliberations, agreement was reached that the Counter Movement Jump and the Squat Jump Test should be used with the GymAware LPT measurement tool for measuring muscle power output at the function level. Since these tests scored equally on the Fibonacci scale, the workgroup selected the Squat Jump Test for the function domain based on its practicality and lower burden on older adults. Additionally, it was agreed that the 30-sec Functional Strength Test primarily measures muscle strength at the functional level rather than muscle power. Given that the COS is focused on assessments of power, the 30-sec Functional Strength Test was excluded. At the activity level, there was consensus that the Timed Up-and-Go, Margaria Kalamen Power Test, Burpee Test, and the 5-times Sit to Stand measure capacity with an emphasis on speed, while the 30-sec Sit to Stand focuses on general capacity, and therefore, was excluded. The Timed Up-and-Go and Margaria Kalamen Power Test received equal scores, however, the workgroup chose the Timed Up-and-Go for its feasibility in an older population. No definitive test or measurement instruments could be agreed upon for the participation domain. However, it was concluded that daily-life accelerometry could be used to monitor habitual physical activity patterns, to see if power training improved the ability to perform activities of daily living. The effectiveness of accelerometry in assessing the impacts of power training remains uncertain, as it has primarily been used to track general physical activity to give insights into intensity and duration. Yet, if coupled with appropriate classification of activities, daily-life accelerometry could provide important insights with regard to the participation domain. The proposed COS to measure the effects of power training in older adults is summarised in Table 5.

Table 5. Core outcome set for measuring the effects of power training in older adults.

ICF health domain	Test
Function	Squat Jump Test
Activities	Timed Up and Go Test
Participation	Daily-life accelerometry

Discussion

Summary of main results

This study is the first to achieve consensus on a core outcome set (COS) of tests to measure the effects of power training in older adults while covering the ICF health domains. The final COS consisted of the Squat Jump Test (paired with accelerometery- based instrumentation) for the function domain and the Timed Up-and-Go Test for the activities domain. These tests were selected according to their ability to measure power output and related activities, making them well-suited for assessing the effects of power training. There was no suitable test identified that could measure the effects of power training within the participation domain of the ICF, but based on the discussion in the expert panel meeting, the use of daily-life accelerometry was suggested as a preliminary measurement instrument until more suitable alternatives are identified.

There has been an ongoing debate regarding the reliability of the Squat Jump Test in the older adult population, as some studies suggest that the Counter Movement Jump may offer superior reliability and sensitivity in the context of pathological conditions, such a sarcopenia (Singh et al., 2014). However, to be sure, future research should compare the clinimetric properties and the feasibility of these tests among the older adult population to determine which is most appropriate for capturing meaningful changes in muscle power. Additionally, it is important to state that an accelerometer is not a specific test, but a device that can be used to measure various activities, causing the

accelerometer to reflect the ICF domain of the test it is paired with. Therefore, the relevance of an accelerometer for measuring muscle power or speed depends on the specific activities assessed. An accelerometer that is used to monitor daily activities (i.e., daily life accelerometry) could be relevant for measuring the effects of power training within the participation domain. Additionally, an accelerometer that is used to estimate power output or speed of performing certain activities, like rising from a chair, may be well-suited for assessing the effects of power training (Regterschot et al., 2015).

While data collection using accelerometry is non-invasive and relatively accessible, the subsequent data analysis is not always straightforward. Accurate interpretation may require knowledge of signal processing and biomechanical modelling to estimate outcomes such as power output. Future research should therefore focus on developing user-friendly and clinically applicable algorithms to enhance the usability of accelerometry in practice. At the same time, it is worth noting that recent technological advancements have made accelerometry increasingly accessible. Modern wearable sensors can now not only quantify activity levels but also estimate power output during specific functional movements, offering new opportunities to assess muscular power in real-world conditions.

The findings reported here are promising, but the current evidence does not yet allow for definitive conclusions regarding the superiority or efficacy of these tests in all settings. Further validation is needed to confirm the utility of the proposed COS, especially in terms of its responsiveness and predictive validity. Therefore, the findings of this study should be interpreted as an important first step in the development of standardised assessment protocols.

Comparison with previous studies

Current literature demonstrates a trend of assessing power training impacts primarily at the body function level of the ICF, focusing on the physiological function of specific muscles, while neglecting how these changes impact everyday activities or social participation (Fragala et al., 2019; Izquierdo et al., 2021). Instead, evaluating the effects of power training and measuring physical functioning in older adults calls for a multifaceted approach (Dobson et al., 2012; Rikli & Jones, 2013). For instance, Alcazar et al. (2018) highlighted the importance of assessing various performance elements to thoroughly evaluate power training effectiveness in older individuals (Alcazar et al., 2018) and Zhang et al. (2024) emphasised the need for ICF-based functional assessment tools specifically tailored for older adults (Zhang et al., 2024). Our study directly addresses this gap by providing a COS of tests specifically structured within the ICF framework, offering a novel contribution to the literature on power training in older adults and supporting a more holistic and clinically meaningful evaluation of training effects. The practical implications of using this framework can be considerable, as it provides clinicians, rehabilitation professionals, and researchers with a standardised set of tools that can guide both assessment and intervention planning. This approach is also in line with O'Neill and Forman (2020) advocacy of a comprehensive test battery to assess physical function as a clinical outcome (O'Neill & Forman, 2020), and is consistent with other studies emphasising muscle strength, physical performance and functional outcomes in older adults participating in power-based training interventions (Alcazar et al., 2018; Byrne et al., 2016; Freiberger et al., 2012).

The lack of standardised outcome measures has been a key limitation in previous research, leading to inconsistencies across studies. Byrne et al. (2016) noted significant variability in outcome measures used in power training studies, making meta-analyses challenging (Byrne et al., 2016). Although standardised effect sizes, such as Cohen's d, can help quantify and compare the magnitude of effects across studies, heterogeneity may be substantial when different instruments are used to assess similar constructs. This comprises the validity of pooled estimates and complicates evidence synthesis. The absence of consensus on which instruments to use still limits interpretability, comparability of constructs, and the ability to inform practice guidelines, making the development of a standardised COS essential. While this study provides an important first step towards standardisation, it does not establish a universally accepted standard. The proposed COS should be seen as a preliminary framework that requires further validation in diverse populations and clinical settings. Future research should focus on evaluating its measurement properties, feasibility in real-world contexts, and alignment with patient-reported outcomes to support broader consensus and implementation.

Our approach to developing a COS set for power training in older adults aligns with and builds upon established methodologies in the field. Williamson et al. (2012) outlined key considerations for developing a COS, emphasising the importance of stakeholder involvement and consensus-building processes (Williamson et al., 2012). We adhered to these principles through the use of a Delphi survey and modified Nominal Group Technique (m-NGT), ensuring diverse expert input and systematic consensus development. Similarly, our methodology resonates with the OMERACT filter 2.0 framework proposed by Boers et al. (2014). This framework

emphasises the need for a COS to be valid and feasible, which aligns with our strategy of categorising tests across different ICF domains and utilising a workgroup to assess practical feasibility of the tests in measuring the effects of power training in older adults (el Hadouchi et al., 2024).

Prinsen et al. (2016) provided guidelines for selecting outcome measurement instruments for a COS (Prinsen et al., 2016). Our study's process of expert consensus aligns with their recommendation for a systematic approach to instrument selection. However, while Prinsen et al. suggest a two-step approach of first defining core outcome domains and then selecting measurement instruments, our study combined these steps in a single consensus process. This modification was made to streamline the process and ensure that selected outcomes were immediately linked to feasible measurement tools.

Strengths and limitations

This study's primary strength lies in its methodological rigour, particularly in achieving consensus among a diverse group of experts. The utilisation of a Delphi survey and the modified Nominal Group Technique (m-NGT) methods facilitated a comprehensive and systematic input from experts, and aligns with best practices in consensus-building research. This approach minimised individual bias and enhanced the applicability of the identified COS. Additionally, the involvement of a working group and collaboration with geriatric physiotherapists ensured that the selected tests would be practical and relevant within the older adult population.

However, several limitations warrant consideration. Despite efforts to include diverse expertise, potential selection bias in expert participation may have influenced outcome prioritisation. The lower response rate for the expert panel meeting compared to the Delphi survey rounds could have impacted the final consensus. Another limitation of this study is its emphasis on performance-based measures, which may not fully reflect the subjective experiences of older adults undergoing power training. This gap underscores the need for further research to combine both objective performance measures and subjective patient experiences, offering a more comprehensive evaluation of the effects of power training in older adults. Future research should also focus on validating these COS in diverse populations and settings to ensure broad applicability and to address the identified limitations. Additionally, rapid advancement of assessment technologies may necessitate periodic updates to the COS, potentially affecting the long-term consistency and comparability of research.

Clinical implications and future research

By incorporating instruments that cover the health domains of the ICF, this study proposes a framework to support a comprehensive evaluation of the impact of power training on the physical functioning of older adults. The proposed COS may help pave the way for more consistent and comparable research and may also support the development of evidence-based practices. This not only enhances the comparability of studies but also contributes to the cumulative knowledge in the field of power training and its broader implications for physical function in older adults. Clinicians, researchers, and practitioners can benefit from a clear framework for evaluating the effectiveness of power training interventions, leading to more comparable and transferable results across studies. This standardisation will also facilitate meta-analyses and systematic reviews, contributing to a more robust evidence base for best practices in power training for older adults.

Future research should focus on validating these COS in diverse populations and settings to ensure their broad applicability. Exploring the integration of new and emerging measurement technologies could further enhance the assessment of power training effects. Additionally, investigating implementation challenges and opportunities associated with adopting these standardised outcomes in clinical and research settings will be crucial for their successful integration into practice.

Conclusions

This study established a consensus-based COS for measuring the effects of power training in older adults. The selected measures, aligned with ICF domains, include the Squat Jump Test (paired with accelerometery-based instrumentation) for the function domain and the Timed Up-and-Go Test for the activities domain. For the participation domain, experts propose using daily-life accelerometry until more suitable outcome measures have been defined and validated.

This comprehensive framework reflects a multifaceted approach to evaluating the impact of power training interventions in older adults. Through standardisation, this COS has the potential to significantly improve the quality, comparability, and clinical relevance of future research. Ultimately, this advancement may lead to more robust evidence-based practices, informing clinical decisions and public health strategies aimed at enhancing

the functional capacity and well-being of older adults through power training interventions.

Acknowledgments

We would like to thank our colleagues at Health Research Consultancy (HRC) and Sports & Science in Lochem, the Netherlands.

Disclosure statement

The primary author holds a dual role as researcher and entrepreneur. To safeguard independence, all data were anonymised, and the Nominal Group Technique session was facilitated by an independent researcher (FM). Delphi participants were academic peers from diverse institutions, and the author did not influence the consensus process.

Funding

This work was supported by the Dutch Research Council (NWO 2017/BOO/00279639).

ORCID

Mohamed El Hadouchi http://orcid.org/0000-0001-9501-6408

References

Alcazar, J., Guadalupe-Grau, A., García-García, F. J., Ara, I., & Alegre, L. M. (2018). Skeletal muscle power measurement in older people: A systematic review of testing protocols and adverse events. *Journals of Gerontology: Series A*, *73*(7), 914–924. https://doi.org/10.1093/gerona/glx216

Al Juffali, L. A., Knapp, P., Al-Aqeel, S., & Watson, M. C. (2019). Medication safety problems priorities in community pharmacy in Saudi Arabia: A multi-stakeholder Delphi study using the human factors framework. *BMJ Open*, *9*(11), e032419. https://doi.org/10.1136/bmjopen-2019-032419

Balachandran, A., Krawczyk, S. N., Potiaumpai, M., & Signorile, J. F. (2014). High-speed circuit training vs hypertrophy training to improve physical function in sarcopenic obese adults: A randomized controlled trial. *Experimental Gerontology*, *60*, 64–71. https://doi.org/10.1016/j.exger.2014.09.016

Bassey, J. E., Fiatarone, M. A., O'neill, E. F., Kelly, M., Evans, W. J., & Lipsitz, L. A. (1992). Leg extensor power and functional performance in very old men and women. *Clinical Science*, *82*(3), 321–327. https://doi.org/10.1042/cs0820321

Bean, J. F., Kiely, D. K., Herman, S., Leveille, S. G., Mizer, K., Frontera, W. R., & Fielding, R. A. (2002). The relationship between leg power and physical performance in Mobility-Limited Older people. *Journal of the American Geriatrics Society*, *50*(3), 461–467. https://doi.org/10.1046/j.1532-5415.2002.50111.x

Bean, J. F., Kiely, D. K., LaRose, S., O'Neill, E., Goldstein, R., & Frontera, W. R. (2009). Increased velocity exercise specific to task training versus the national institute on aging's strength training program: Changes in limb power and mobility. *Journals of Gerontology - Series A Biological Sciences & Medical Sciences*, *64*(9), 983–991. https://doi.org/10.1093/gerona/glp056

Boers, M., Kirwan, J. R., Wells, G., Beaton, D., Gossec, L., d'Agostino, M.-A., Conaghan, P. G., Bingham, C. O., Brooks, P., Landewé, R., March, L., Simon, L. S., Singh, J. A., Strand, V., & Tugwell, P. (2014). Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *Journal of Clinical Epidemiology*, *67*(7), 745–753. https://doi.org/10.1016/j.jclinepi.2013.11.013

Byrne, C., Faure, C., Keene, D. J., & Lamb, S. E. (2016). Ageing, muscle power and physical function: A systematic review and implications for pragmatic training interventions. *Sports Medicine*, *46*(9), 1311–1332. https://doi.org/10.1007/s40279-016-0489-x

Cadore, E. L., Rodríguez-Mañas, L., Sinclair, A., & Izquierdo, M. (2013). Effects of different exercise interventions on risk of falls, gait ability, and balance in physically frail older adults: A systematic review. *Rejuvenation Research*, *16*(2), 105–114. https://doi.org/10.1089/rej.2012.1397

Chiarotto, A., Terwee, C. B., Deyo, R. A., Boers, M., Lin, C.-W. C., Buchbinder, R., Corbin, T. P., Costa, L. O., Foster, N. E., Grotle, M., Koes, B. W., Kovacs, F. M., Maher, C. G., Pearson, A. M., Peul, W. C., Schoene, M. L., Turk, D. C., van Tulder, M. W., & Ostelo, R. W. (2014). A core outcome set for clinical trials on non-specific low back pain: Study protocol for the development of a core domain set. *Trials*, *15*(1), 511. https://doi.org/10.1186/1745-6215-15-511

Delmonico, M. J., Harris, T. B., Visser, M., Park, S. W., Conroy, M. B., Velasquez-Mieyer, P., Boudreau, R., Manini, T. M., Nevitt, M., Newman, A. B., & Goodpaster, B. H. (2009). Longitudinal study of muscle strength, quality, and adipose tissue infiltration. *The American Journal of Clinical Nutrition*, *90*(6), 1579–1585. https://doi.org/10.3945/ajcn.2009.28047

Dobson, F., Hinman, R. S., Hall, M., Terwee, C. B., Roos, E. M., & Bennell, K. L. (2012). Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: A systematic review. *Osteoarthritis and Cartilage*, *20*(12), 1548–1562. https://doi.org/10.1016/j.joca.2012.08.015

Ferrucci, L., Cooper, R., Shardell, M., Simonsick, E. M., Schrack, J. A., & Kuh, D. (2016). Age-related change in mobility: Perspectives from life course epidemiology and Geroscience. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *71*(9), 1184–1194. https://doi.org/10.1093/gerona/glw043

Fielding, R. A., Vellas, B., Evans, W. J., Bhasin, S., Morley, J. E., Newman, A. B., Abellan van Kan, G., Andrieu, S., Bauer, J., Breuille, D., Cederholm, T., Chandler, J., De Meynard, C., Donini, L., Harris, T., Kannt, A., Keime Guibert, F., Onder, G., Papanicolaou, D. … Zamboni, M. (2011). Sarcopenia: An undiagnosed condition in older adults. Current consensus definition: Prevalence, etiology, and consequences. International working group on sarcopenia. *Journal of the American Medical Directors Association*, *12*(4), 249–256. https://doi.org/10.1016/j.jamda.2011.01.003

Foldvari, M., Clark, M., Laviolette, L. C., Bernstein, M. A., Kaliton, D., Castaneda, C., Pu, C. T., Hausdorff, J. M., Fielding, R. A., & Singh, M. A. F. (2000). Association of muscle

power with functional status in community-dwelling elderly women. *Journal of Gerontology*, *55*(4), 192–199. https://doi.org/10.1093/gerona/55.4.m192

Fragala, M. S., Cadore, E. L., Dorgo, S., Izquierdo, M., Kraemer, W. J., Peterson, M. D., & Ryan, E. D. (2019). Resistance training for older adults: Position statement from the national strength and conditioning association. *Journal of Strength and Conditioning Research*, *33*(8), 2019–2052. https://doi.org/10.1519/JSC.0000000000003230

Freiberger, E., de Vreede, P., Schoene, D., Rydwik, E., Mueller, V., Frändin, K., & Hopman-Rock, M. (2012). Performance-based physical function in older community-dwelling persons: A systematic review of instruments. *Age and Ageing*, *41*(6), 712–721. https://doi.org/10.1093/ageing/afs099

Gagnier, J. J., Page, M. J., Huang, H., Verhagen, A. P., & Buchbinder, R. (2017). Creation of a core outcome set for clinical trials of people with shoulder pain: A study protocol. *Trials*, *18*(1), 336. https://doi.org/10.1186/s13063-017-2054-9

Hadouchi, M., Kiers, E., Boerstra, B., Berends, R., Schreuder, E., Veenhof, C., & van Dieën, J. (2024). The reliability, validity, and responsiveness of tests used to assess the effects of power training in older adults: A systematic review. *Physical Therapy Reviews*, *29*(4), 1–14. https://doi.org/10.1080/10833196.2024.2376439

Hadouchi, M., Kiers, H., de Vries, R., Veenhof, C., & van Dieën, J. (2022). Effectiveness of power training compared to strength training in older adults: A systematic review and meta-analysis. *European Review of Aging and Physical Activity*, *19*(1). https://doi.org/10.1186/s11556-022-00297-x

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Humphrey-Murto, S., Varpio, L., Gonsalves, C., & Wood, T. J. (2017). Using consensus group methods such as delphi and nominal group in medical education research. *Medical Teacher*, *39*(1), 14–19. https://doi.org/10.1080/0142159X.2017.1245856

Izquierdo, M., Merchant, R. A., Morley, J. E., Anker, S. D., Aprahamian, I., Arai, H., Aubertin-Leheudre, M., Bernabei, R., Cadore, E. L., Cesari, M., Chen, L.-K., de Souto Barreto, P., Duque, G., Ferrucci, L., Fielding, R. A., García-Hermoso, A., Gutiérrez-Robledo, L. M., Harridge, S. D. R., & Vellas, B. (2021). International exercise recommendations in older adults (ICFSR): Expert consensus guidelines. *The Journal of nutrition, Health & Aging*, *25*(7), 824–853. https://doi.org/10.1007/s12603-021-1665-8

Jones, J., & Hunter, D. (1995). Qualitative research: Consensus methods for medical and health services research. *BMJ*, *311* (7001), 376–380. https://doi.org/10.1136/bmj.311.7001.376

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of Advanced Nursing*, *53*(2), 205–212. https://doi.org/10.1111/j.1365-2648.2006.03716.x

Kraemer, W., Denegar, C., & Flanagan, S. (2009). Recovery from injury in sport: Considerations in the transition from medical care to performance care. *Sports Health: A Multidisciplinary Approach*, *1*(5), 392–395. https://doi.org/10.1177/1941738109343156

McMillan, S. S., King, M., & Tully, M. P. (2016). How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*. https://doi.org/10.1007/s11096-016-0257-x

Metter, E. J., Conwit, R., Tobin, J., & Fozard, J. L. (1997). Age-associated loss of power and strength in the upper extremities in women and men. *Journal of Gerontology: Biological Sciences*, *52*(5), B267–B276. https://doi.org/10.1093/gerona/52a.5.b267

Miszko, T. A., Cress, M. E., Slade, J. M., Covey, C. J., Agrawal, S. K., & Doerr, C. E. (2003). Effect of strength and power training on physical function in community-dwelling older adults. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *58*(2), M171–M175. https://doi.org/10.1093/gerona/58.2.M171

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

O'Neill, D., & Forman, D. E. (2020). The importance of physical function as a clinical outcome: Assessment and enhancement. *Clinical Cardiology*, *43*(2), 108–117. https://doi.org/10.1002/clc.23311

Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - A practical guideline. *Trials*, *17*(1), 449. https://doi.org/10.1186/s13063-016-1555-2

Regterschot, G. R. H., Zhang, W., Baldus, H., Stevens, M., & Zijlstra, W. (2015). Sensor-based monitoring of sit-to-stand performance is indicative of objective and self-reported aspects of functional status in older adults. *Gait & Posture*, *41*(4), 935–940. https://doi.org/10.1016/j.gaitpost.2015.03.350

Rikli, R. E., & Jones, C. J. (2013). Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. *The Gerontologist*, *53*(2), 255–267. https://doi.org/10.1093/geront/gns071

Signorile, J. F., Carmel, M. P., Lai, S., & Roos, B. A. (2005). Early plateaus of power and torque gains during high- and low-speed resistance training of older women. *Journal of Applied Physiology*, *98*(4), 1213–1220. https://doi.org/10.1152/japplphysiol.00742.2004

Singh, H., Kim, D., Kim, E., Bemben, M. G., Anderson, M., Seo, D.-I., & Bemben, D. A. (2014). Jump test performance and Sarcopenia status in men and women, 55 to 75 years of age. *Journal of Geriatric Physical Therapy*, *37*(2), 76–82. https://doi.org/10.1519/JPT.0b013e3182a51b11

Skelton, D. A., Young, A., Greig, C. A., & Malbut, K. E. (1995). Effects of resistance training on strength, power, and selected functional abilities of women aged 75 and older. *Journal of the American Geriatrics Society*, *43*(10), 1081–1087. https://doi.org/10.1111/j.1532-5415.1995.tb07004.x

Steib, S., Schoene, D., & Pfeifer, K. (2010). Dose-response relationship of resistance training in older adults. *Medicine & Science in Sports & Exercise*, *42*(5), 902–914. https://doi.org/10.1249/MSS.0b013e3181c34465

Strotmeyer, E. S., Winger, M. E., Cauley, J. A., Boudreau, R. M., Cusick, D., Collins, R. F., Chalhoub, D., Buehring, B., Orwoll, E., Harris, T. B., & Caserotti, P. (2018). Normative values of

muscle power using force plate Jump tests in men aged 77-101 years: The osteoporotic fractures in men (MrOS) study. *The Journal of nutrition, Health & Aging*, *22*(10), 1167–1175. https://doi.org/10.1007/s12603-018-1081-x

Tieland, M., Trouwborst, I., & Clark, B. C. (2018). Skeletal muscle performance and ageing. *Journal of Cachexia, Sarcopenia and Muscle*, *9*(1), 3–19. https://doi.org/10.1002/jcsm.12238

Tschopp, M., Sattelmayer, M. K., & Hilfiker, R. (2011). Is power training or conventional resistance training better for function in elderly persons? A meta-analysis. *Age and Ageing*, *40*(5), 549–556. https://doi.org/10.1093/ageing/afr005

Vergouw, D., Heymans, M. W., de Vet, H. C., van der Windt, D. A., & van der Horst, H. E. (2011). Prediction of persistent shoulder pain in general practice: Comparing clinical consensus from a Delphi procedure with a statistical scoring system. *BMC Family Practice*, *12*(1), 63. https://doi.org/10.1186/1471-2296-12-63

Vogel, C., Zwolinsky, S., Griffiths, C., Hobbs, M., Henderson, E., & Wilkins, E. (2019). A delphi study to build consensus on the definition and use of big data in obesity research. *International Journal of Obesity*, *43*(12), 2573–2586. https://doi.org/10.1038/s41366-018-0313-9

Wiegmann, S., Felsenberg, D., Armbrecht, G., & Dietzel, R. (2021). Longitudinal changes in muscle power compared to muscle strength and mass. *Journal of Musculoskeletal and Neuronal Interactions*, *21*(1), 13–25.

Williamson, P. R., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., & Tugwell, P. (2012). Developing core outcome sets for clinical trials: Issues to consider. *Trials*, *13*(1), 132. https://doi.org/10.1186/1745-6215-13-132

World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. World Health Organization. https://iris.who.int/handle/10665/42407

World Health Organization. (2017). *Decade of healthy ageing 2020–2030*. WHO.

World Health Organization. (2017). *Global strategy and action plan on ageing and health*. World Health Organization.

Zhang, M., Chen, L., Dai, J., Yang, Q., Huang, Z., He, J., Ji, H., Sun, J., & Li, D. (2024). Application of a new monitoring variable: Effects of power loss during Squat training on strength gains and sports performance. *Journal of Strength & Conditioning Research*, *38*(4), 656–670. https://doi.org/10.1519/JSC.0000000000004677

Copyright of Journal of Sports Sciences is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.