Mind your step

Citation for published version (APA):

Gafner, S. C. (2022). *Mind your step: The use of hip abductor strength to assess the fall risk of persons 65 years and older*. [Doctoral Thesis, Maastricht University]. Gildeprint Drukkerijen.
<https://doi.org/10.26481/dis.20220114sg>

Document status and date:

Published: 01/01/2022

DOI:

[10.26481/dis.20220114sg](https://doi.org/10.26481/dis.20220114sg)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

MIND YOUR STEP

The use of hip abductor strength to assess the fall
risk of persons 65 years and older

Simone Chantal Gafner

MIND YOUR STEP

The use of hip abductor strength to assess
the fall risk of persons 65 years and older

DISSERTATION

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof.dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public on the 14th of January 2022, at 14hours

by
Simone Chantal Gafner

Author: Simone Chantal Gafner
Printing: Gildeprint Enschede, gildeprint.nl
Layout and cover design: Birgit Vredenburg, persoonlijkproefschrift.nl
This cover has been designed using resources from [Freepik.com](http://freepik.com)

ISBN: 978-94-6419-369-5
Copyright 2021 © Simone Gafner
All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval
system or transmitted in any form or by any means without permission of the author.

Supervisor(s):

Prof dr RA de Bie (UM)
Dr CHG Bastiaenen (UM)

Co-supervisor:

Prof dr Lara de Preux-Allet (School of Health Sciences, HES-SO Valais-Wallis,
University of Applied Sciences and Arts Western Switzerland, Valais)

Assessment Committee:

Prof dr JMGA Schols (chair) (UM)
Prof dr ED de Bruin (Karolinske Institute, Stockholm)
Dr JSM Hobbelen (Hanze University of Applied Sciences, Groningen)
Prof dr AF Lenssen (UM)
Dr C McCrum (UM)

CONTENTS

Preface	6
Chapter 1	General Introduction 9
Chapter 2	Evaluation of hip abductor and adductor strength in the elderly: a reliability study 25
Chapter 3	Intra-rater reliability of hip abductor isometric strength testing in a standing position in older fallers and non-fallers 43
Chapter 4	Hip muscle and hand grip strength to differentiate between older fallers and non-fallers; a cross-sectional validity study 65
Chapter 5	The role of hip abductor strength in identifying older persons at risk of falls: A diagnostic accuracy study 85
Chapter 6	Reliability and validity of an adapted hip abductor strength measure in a closed chain position. A new fall risk assessment for older persons? A study protocol 105
Chapter 7	Reliability and diagnostic accuracy of commonly used performance tests relative to fall history in older persons: a systematic review 131
Chapter 8	General Discussion 177
Chapter 9	Impact Paragraph 199
Addenda	205
	Summary 206
	Zusammenfassung 212
	Résumé 220
	Acknowledgments 228
	About the author 234

PREFACE

Imagine it is a sunny Sunday afternoon in August. It is your grandmother's birthday, and your family is gathered together in her garden. Your grandmother is turning 89 years old today. She has been living alone in a big house since your grandfather died 5 years ago.

As you are enjoying the warm sun on your face, you look at the scene around you and are pleased to see all these happy faces. Your father (who retired 10 years ago) is standing in front of the barbecue. Your grandmother is walking arm in arm with your mother (who is also retired), and they approach the grill slowly. Their walk looks a bit like a dance with their left to right deviating steps through the grass. Once they reach the grill, your father places some meat on your grandmother's plate. Your mother adds some salad that your grandmother is choosing from the buffet nearby. Together, they walk their little dance-like walk back to the nicely decorated table.

You are sitting in your chair, pretty full after all the delicious food you ate. You are feeling great and happy to share this moment with your loved ones. Everything is perfect in that moment, and you think to yourself:

Everything should stay like this.

You wish for your parents and your grandmother to stay autonomous for as long as possible. However, one big risk for persons who are 65 years and older is that of falling. You know about the possible consequences associated with a fall, which amongst other injuries, can lead to dependency and an increased mortality.

At that moment, consider the following:

- What would you be willing to do to return to this garden in a year and have a birthday party with the same people present who are in the same conditions of autonomy?
- What would you like to know to support your loved ones in preventing a fall?
- Which questions would you like to have answered in relation to fall risk?
 - *Which factors influence falls?*
 - *What tests can evaluate those fall-influencing factors?*
 - *What does science say regarding fall risk and fall-risk assessments?*

After outlining this nice garden scenery with its personal and emotional aspects, the following dissertation takes a scientific approach to this topic and will answer some of the above questions.

CHAPTER 1

General Introduction

Demographic changes are leading to a drastically increasing number of people who are 65 years and older, which is the world's fastest growing age group (1). In their World Populations Prospects (1), the United Nations states that one in six people in the world and one in four people living in Europe and North America will be older than 65 by 2050 (1). They also project that the number of people over age 80 will triple, from 143 million in 2019 to 426 million in 2050 (1). This implicates that physical therapists and other health care professionals will have an increasing number of patients who are 65 years and older in their daily practice in the near future.

FALLS AND FALL-INFLUENCING FACTORS

One of ageing's important complications is the increasing risk of falls. About one-third of community dwellings and over half of the people aged 65 years and older in long-term care institutions fall at least once a year (2). Those impressive numbers of the high frequency of falls, together with the demographic changes, show that fall risk detection and prevention is a global public health challenge. The high fall rate is particularly disconcerting due to the serious consequences of falls, such as a loss of independence (3), fear of falling (4), depression (4) and increased mortality (5). Additionally, in terms of the actual fall, there is a high risk that older persons will enter in a vicious cycle of deconditioning, which could lead to future falls due to increased fear of falling again and the related reduced activity (6).

The etiology of falls is multifactorial and can be categorized in different groups such as environmental factors, psychological factors, cognitive factors, factors linked to medication and physiological factors (7). Therefore, many different factors influence the "fall risk" construct as it is used in this thesis. However, the focus will be on the physiological factors, the parameters that are modifiable and important for guiding the content of a physical therapy intervention, such as decreased balance, strength and mobility. In fact, functional lower-extremity weakness in older persons is a strong indicator that could enhance the unfortunate transition from a non-faller into a faller (8), and its assessment and treatment for persons at risk of falls is recommended (9). This thesis focused to a greater extent on lower limb strength because its assessment might be an interesting point of departure for physical therapists and other health care professionals to evaluate their patients' fall risk and developing insight in a directly targetable parameter within the context of an intervention.

Additionally, to the fall related consequences on a personal level, falls also represent an important economic burden (10), which is likely greater than policymakers recognize (11). In Switzerland, about 87,000 persons who are 65 and older experience a fall-related injury per year, and about 1,500 persons die due to a fall (12).

Although the costs that falls cause are important on an economic level and may give important information about the problem's vastness, it might also indicate the potential for effective prevention strategies. Nevertheless, the personal fate of the considered persons seems to be at least as important as the economic burden.

Health care professionals should strive to detect persons at risk of falls as soon as possible, to be able to shape an intervention integrating the patient's needs and wishes and, if possible, to prevent a first fall or thereafter prevent or at least postpone recurrent falls. To do this, "easy to implement" clinical fall-risk assessment tests with a high diagnostic accuracy, combined with a low burden of a quick and easy patient assessment, are essential.

WHAT TESTS ARE THERE FOR EVALUATING FALL-INFLUENCING FACTORS?

Early detection of an increased risk of falling is crucial for the at-risk population, as is providing health care professionals with information for a targeted fall prevention and rehabilitation. The great diversity of existing fall-risk assessments reflects the topic's importance, but the number of options makes it difficult for the health care professionals to choose the appropriate test.

There are several risk factors for falls and many fall-risk assessment tests have been developed (13). Among these, physical performance tests such as lower limb strength assessments, balance performance tests and gait speed assessment are important as they play a role in activities such as standing up, walking or turning around (14-16).

Within the field of performance tests, fall-risk assessment tests are divided into single or multidimensional performance tests, which are usually included in an extensive test battery. The following two paragraphs provide a few examples of existing tests.

Commonly used single performance tests include gait speed (17), single leg stance (the amount of time that an older person can stand on one leg without arm support) (18, 19) and a functional reach test (the maximum reach distance beyond an arm's length while maintaining a fixed base of support in the standing position) (20).

Commonly used multidimensional performance tests include the Performance Oriented Mobility Assessment (POMA) (21), the Timed Up and Go (TUG) test (15) and the Short Physical Performance Battery (SPPB) (22). POMA, also known as the Tinetti balance and gait scale, is one of the earliest tests created to assess balance, gait and fall risk in older persons (23-25). It includes evaluating balance under perturbed conditions as well as evaluating gait characteristics. Lower scores on POMA are associated with an increased fall risk (24). The TUG test assesses persons' mobility

(static and dynamic balance) (15, 23). Recent guidelines suggest using the TUG in clinical settings to assess fall risk (23). The SPPB is a functional physical measure of performance for older persons (26) consisting of three sub-scores assessing balance performance, walking speed and lower limb strength (27-29).

Given the large amount of existing fall-risk-assessment tests for everyday clinical use (e.g., 64 identified fall-risk assessment tests in a systematic review of Lusardi et al. (13)), the choice is not always easy for the “best” diagnostic tool or combination of diagnostic tools in relation to a pre-specified individual goal. Therefore, it is essential to know and interpret aspects of such tests’ diagnostic accuracy in various conditions to support health care professionals to choose the most appropriate test (or battery of tests) for each individual member of the target population and setting.

IMPORTANT PARAMETERS FOR THE CHOICE OF A TEST FOR FALL-RISK ASSESSMENT

The appropriate fall-risk assessment within a diagnostic process should provide insight in patients’ individual needs and goals, and ideally indicate for an individually applied, well-designed intervention for prevention or rehabilitation. Alongside a measurement instrument’s quality, practicability aspects will be important for the health care professional. As such, we are interested in the duration of the test administration and the needed material, which are both important features to judge the feasibility in the specific setting.

The Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) group defines the quality of a measurement instrument as depending on reliability, validity and responsiveness (30). Thus, we will shortly explain the theoretical aspects of reliability and validity.

Reliability

Reliability, in the context of this thesis, is “the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g., over time (test–retest); by different raters on the same occasion (inter-rater); or by the same raters (i.e., raters or responders) on different occasions (intra-rater)” (31). Reliability can be expressed in absolute and/or relative reliability (32). Relative reliability describes to what degree individuals maintain their position in a sample with repeated measurements (32). In other words, it concerns the ratio of the individual variance divided by the total variance (33). Absolute reliability is an individual’s degree of variation between repeated measures (32). Thus, it quantifies the scores’ reliability within an individual participant on different occasions (34).

Relative reliability can be expressed/calculated in more ways and provides information about how well a measurement can distinguish between persons (35). The COSMIN group is recommending calculating ICCs for continuous variables (33). In this thesis, we are interested in the relative reliability between two measurements (intra-rater) or between two raters (inter-rater) and will thus use the ICC. The variation/heterogeneity in the population is a parameter that is highly influencing reliability (33, 35). Therefore, reliability is only generalizable for a sample with similar variation as the one in which it was initially evaluated.

Absolute reliability is often expressed as the standard error of measurement (SEM), which is an indicator for agreement (35), Bland and Altman plots and/or limits of agreement (LoA) (36).

Validity

Validity of a measurement instrument indicates the degree to which it measures the construct it purports to measure (30). According to the COSMIN group, validity can be subdivided into three parts: content, construct and criterion validity (33). The COSMIN panel defined content validity as “the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured” (31). Out of the many different factors influencing the “fall risk” construct (7), this thesis focuses on modifiable risk factors potentially prone to a future intervention. Falls often occur during physical activities such as walking or climbing stairs (8). Therefore, part of the fall risk construct can be assessed with physical performance tests such as lower limb strength assessments, balance performance tests or gait speed assessment. In the context of persons 65 years and older, it is important that the test is feasible and the assessment time be relatively short.

The construct validity is usually assessed when a criterion standard is missing and is based on the assumption that the new measurement instrument validly assesses the construct it is supposed to measure (33).

The criterion validity is defined as follows: “The degree to which the scores of a measurement instrument are an adequate reflection of a criterion standard” (31). It has two subdivisions: a) the concurrent validity assessing the score for the measurement as well as the criterion standard and b) the predictive validity assessing whether the measurement predicts the criterion standard in the future (33). The diagnostic accuracy of a measurement instrument (sensitivity, specificity, positive and negative predictive values [PPV and NPV], positive and negative likelihood ratios [LR+ and LR–], and positive and negative post-test probability) is considered part of the criterion validity. When clinicians have to choose a measurement instrument, the intrinsic ability of a test to detect persons at risk (true positives, sensitivity) or not

at risk of falls (true negatives, specificity) are important. Once the result is known, the intrinsic qualities are of less interest and the individual level of the outcome becomes more important (PPV, NPV, LR+, LR– and positive and negative post-test probabilities). On a personal level, the PPV and NPV answer the question, “If my test result is positive, what are the chances that I really am at risk of falls?” (37). Regarding the likelihood ratios (LRs); LR+ indicates how much more likely a positive test result is in people with the disease/disorder than it is in people without the disease/disorder (13, 37), and LR– indicates how much less likely a negative test result is in people with the disease/disorder than it is in people without the disease (13, 37). The LR is calculated as a ratio based on sensitivity and specificity and, except from spectrum bias, independent from prevalence (37). The larger the value above 1 for LR+ or the smaller the value below 1 for LR–, the more valuable the test result is (37). The post-test probabilities are highly important for researchers and clinicians, but they are not yet well understood in the clinical setting where the test will be implemented. Post-test probabilities help to interpret the test results together with the patient and should, together with the intrinsic qualities and other diagnostic accuracy values, influence the choice of the fall-risk assessment. The post-test probability is an indication of the shift of certainty of the fall risk of a person after performance of the test (13). Post-test probability can be calculated with the help of positive or negative LRs (LR+ or LR–) (13) and is dependent on the prevalence of the fall risk in the tested population.

Aspects to consider when choosing the best adapted measurement instrument

From a health care professional's perspective, the person at risk and their need to prevent a fall should be the center of focus, and individual goal setting should be established together with the patient. To do so, the following aspects of a measurement instrument that should be considered to obtain good insights into individual needs can be summarized as follows:

- 1) The purpose for which the test is administered (evaluation of a treatment, for diagnostic reasons or for prediction of future events) should be clearly defined (33).
- 2) Feasibility and good reliability of a measurement instrument are a prerequisite for the validity and express the degree to which a measurement tool is free from measurement error (30).
- 3) When a measurement instrument must be chosen, the test's intrinsic qualities for detecting true positives and true negatives (sensitivity and specificity) are of interest.

- 4) Once the tests are performed, the focus shifts and the tested person is primarily interested in knowing the chance that a clinician is giving them the right answer. For that, the intrinsic qualities of a test are less important, and the health care professional should focus much more on values that give information about the probability that a certain result of a specific test is right or wrong (PPV, NPV, LR+, LR– and positive and negative post-test probabilities).

As described in the *Falls and Fall-Influencing Factors* section, falls have multifactorial etiology (7). In the context of my background as a physiotherapist, the focus of this thesis is on performance-based measures because they assess modifiable risk factors. Knowing the influence of lower limb strength on fall risk, I was interested in investigating a new, potentially interesting factor that might influence fall risk and could be used in addition to the currently used performance-based measurements.

LOWER LIMB MUSCLE STRENGTH AND RISK OF FALLS

As mentioned, muscle strength is important in preventing falls (38), and functional lower-extremity weakness in older persons is assumed to enhance their transition from being a non-faller to a faller (8). Reduced muscle strength further increases the risk of mobility limitations (39-41).

Most studies about lower limb strength in relation to fall risk have been performed on muscles moving parts of the lower limb in the sagittal plane (flexion and extension of the hip, knee and ankle) (42). Only a few investigated the influence of muscle groups mostly acting in the frontal plane in relation to the fall risk of older adults (43-45).

Greater strength of the hip muscle groups seemed to influence components of balance recovery among older persons (longer step length and decreased reaction time) (45). Muscle strength in the hip frontal plane and spinal moments of force are the dominant muscles to counteract destabilization in the mediolateral direction during single-leg support and to stabilize the head, arms and trunk over the support leg (46). Thus, these factors prevent the center of mass from rapidly falling downwards and laterally toward the unsupported swing side (46), therefore causing one to lose balance. Thus, it seems that interest in hip muscle strength as a possible contributing factor to the fall risk of older people is increasing, having led to studies in this field (43, 44, 47). Consequently, those facts led me to a closer investigation of the influence of hip muscle strength, specifically frontal plane muscle strength, on the risk of falls among older persons.

HIP ABDUCTOR STRENGTH

Hip abductor muscle strength in older persons is assumed to be important for several reasons.

Recent literature underlines the significance of hip abductor muscle strength to maintain mediolateral balance control (48-50), a factor that is often reduced in older persons, especially fallers (44), and assumed to lead to lateral and posterolateral falls (51). This is of clinical relevance because posterolateral falls have high hip injury potential and lead to fractures (52-54). Furthermore, as mentioned above, hip frontal plane muscle strength and spinal moments of force are the dominant muscles to counteract destabilization in the mediolateral direction during single leg support (46). Moreover, hip abductor muscles, a muscle group mostly functioning as stabilizers, showed a more pronounced decline in strength compared to hip extensor muscle in older persons (55), a factor that might also lead to this decreased mediolateral stability cited above, as well as an increased risk of falls. Lastly, hip abductor muscle contractions at the moment of impact after a fall seemed to act protectively against the stress induced on the femoral neck and likewise the risk of a femoral fracture (56).

The mentioned facts led to the hypothesis that hip abductor strength could be a promising contributing factor for assessing fall risk. Because this approach seemed to be clinically relevant and novel, we decided to investigate the hip abductor muscle group in greater detail and, if our hypotheses of its importance can be proven, planned to establishment cut-off values and diagnostic accuracy to facilitate the clinical decision-making process regarding older individuals' fall risks. The validation of this approach provides clinicians not only an additional fall-risk assessment to the already existing test battery but might also give crucial information about a directly targetable parameter for fall-risk assessment. We will hypothesize for the future that if hip frontal plane muscle strength measures can be validated as clinically important fall risk predictors, then the benefits will concern a substantial percentage (30-50% according to the WHO (2)) of the permanently growing population of people older than 65 years.

THESIS AIMS

The general aim of this thesis is to gain insight into the role of hip abductor muscle strength in the field of fall risk in older adults.

To do so, we addressed two main issues in Part 1 and Part 2 of this thesis.

Part 1: The role of hip abductor strength in the field of fall-risk assessment and development of a new measurement instrument and procedure for daily clinical use

It was planned to assess feasibility, reliability and validity of our measurement method for hip frontal plane strength assessments. In a later emerging project, we further develop a hip abductor strength test to obtain a new and functional testing procedure that is easy to use. As such, for Part 1, we aimed for the following sub-sequential steps:

- 1) To obtain insight into the measurement method in question and important information for the following examination of its validity, our **first aim** was to investigate the reliability and feasibility of our hip abductor and adductor strength test in the side-lying position with a dynamometer fixed to a custom-made frame (Chapter 2).
- 2) The **second aim** was to assess the reliability and feasibility of a more functional and less time-consuming assessment of hip abductor strength in a standing position in persons aged 65 years and older (Chapter 3).
- 3) The **third aim** was to assess the discriminant ability of all hip muscles groups to differentiate between older fallers and non-fallers. One of the underlying goals of this step was, as a novel approach, to investigate our hypothesis and confirm whether hip abductor strength has the ability to distinguish fallers and non-fallers as well as to examine this capacity in relation to all the other hip muscle groups in the same study (Chapter 4).
- 4) To allow the use of hip abductor strength to discriminate between fallers and non-fallers in everyday clinical practice, we evaluated the cut-off values and the measurement's sensitivity, specificity, positive and negative predictive values, and the positive and negative LR's as our **fourth aim** (Chapter 5).

Because of the results and experiences of our first four projects, an additional project emerged and was added to this dissertation.

- 5) The development of a new measurement device and procedure allowing an easy and quick assessment of hip abductor strength in a standing, closed-chain position became our **fifth aim**. The study protocol for the investigation of its reliability and validity in a future study can be found in Chapter 6.

Part 2: The diagnostic accuracy of currently used fall-risk assessments (including strength measures) to gain insight into their ability for fall risk detection and/or prediction

Part 2 was based on the fact that falls have a multifactorial etiology and the awareness that assessing hip abductor strength thus must be combined with other functional fall-risk assessments to increase the diagnostic accuracy for the detection of persons at risk of falls. Consequently, we defined the following aim for Part 2.

- 6) To perform a systematic review of currently used fall-risk assessments to provide evidence that helps health care professionals choose the most suitable test for their clinical use (Chapter 7; **Aim 6**).

REFERENCES

1. United Nations DoEaSA, Population Division (2019). United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423).
2. WHO World Health Organization report/ WHO Global report on falls prevention in older age. 2007. ISBN 978 92 4 156353 6, 2007.
3. Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC geriatrics*. 2011;11:33.
4. Iaboni A, Flint AJ. The complex interplay of depression and falls in older adults: a clinical review. *Am J Geriatr Psychiatry*. 2013;21(5):484-92.
5. Liu SW, Obermeyer Z, Chang Y, Shankar KN. Frequency of ED revisits and death among older adults after a fall. *The American journal of emergency medicine*. 2015;33(8):1012-8.
6. Cumming RG, Salkeld G, Thomas M, Szonyi G. Prospective study of the impact of fear of falling on activities of daily living, SF-36 scores, and nursing home admission. *The journals of gerontology Series A, Biological sciences and medical sciences*. 2000;55(5):M299-305.
7. Sousa LM, Marques-Vieira CM, Caldevilla MN, Henriques CM, Severino SS, Caldeira SM. Risk for falls among community-dwelling older people: systematic literature review. *Revista gaucha de enfermagem*. 2017;37(4):e55030.
8. Muir SW, Berg K, Chesworth BM, Klar N, Speechley M. Modifiable Risk Factors Identify People Who Transition from Non-fallers to Fallers in Community-Dwelling Older Adults: A Prospective Study. *Physiotherapy Canada Physiotherapie Canada*. 2010;62(4):358-67.
9. Panel on Prevention of Falls in Older Persons AGS, British Geriatrics S. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. *Journal of the American Geriatrics Society*. 2011;59(1):148-57.
10. Heinrich S, Rapp K, Rissmann U, Becker C, König HH. Cost of falls in old age: a systematic review. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2010;21(6):891-902.
11. Davis JC, Robertson MC, Ashe MC, Liu-Ambrose T, Khan KM, Marra CA. International comparison of cost of falls in older adults living in the community: a systematic review. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2010;21(8):1295-306.
12. Beratungsstelle für Unfallverhütung BFU. Status 2019: Statistik der Nichtberufsunfälle und des Sicherheitsniveaus in der Schweiz. Strassenverkehr, Sport, Haus und Freizeit. BFU. 2019.

13. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E, et al. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. *Journal of geriatric physical therapy* (2001). 2017;40(1):1-36.
14. Thapa PB, Brockman KG, Gideon P, Fought RL, Ray WA. Injurious falls in nonambulatory nursing home residents: a comparative study of circumstances, incidence, and risk factors. *Journal of the American Geriatrics Society*. 1996;44(3):273-8.
15. Schoene D, Wu SM, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, et al. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. *Journal of the American Geriatrics Society*. 2013;61(2):202-8.
16. Nelson AJ, Certo LJ, Lembo LS, Lopez DA, Manfredonia EF, Vanichpong SK, et al. The functional ambulation performance of elderly fallers and non-fallers walking at their preferred velocity. *NeuroRehabilitation*. 1999;13(3):141-6.
17. Veronese N, Bolzetta F, Toffanello ED, Zambon S, De Rui M, Perissinotto E, et al. Association between Short Physical Performance Battery and falls in older people: the Progetto Veneto Anziani Study. *Rejuvenation research*. 2014;17(3):276-84.
18. Lundin H, Saaf M, Strender LE, Nyren S, Johansson SE, Salminen H. One-leg standing time and hip-fracture prediction. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2014;25(4):1305-11.
19. Chang CJ, Chang YS, Yang SW. Using single leg standing time to predict the fall risk in elderly. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference*. 2013;2013:7456-8.
20. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. *Journal of gerontology*. 1990;45(6):M192-7.
21. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *The New England journal of medicine*. 1988;319(26):1701-7.
22. 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. *Journal of gerontology*. 1994;49(2):M85-94.
23. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: a review of the literature. *Maturitas*. 2013;75(1):51-61.
24. Alexander NB, Goldberg A. Clinical gait and stepping performance measures in older adults. *European Review of Aging and Physical Activity*. 2006;3(1):20.
25. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *Journal of the American Geriatrics Society*. 1986;34(2):119-26.
26. Freiburger E, de Vreede P, Schoene D, Rydwick E, Mueller V, Frandin K, et al. Performance-based physical function in older community-dwelling persons: a systematic review of instruments. *Age and ageing*. 2012;41(6):712-21.
27. Tanimoto Y, Watanabe M, Sun W, Sugiura Y, Hayashida I, Kusabiraki T, et al. Sarcopenia and falls in community-dwelling elderly subjects in Japan: Defining sarcopenia according to criteria of the European Working Group on Sarcopenia in Older People. *Archives of gerontology and geriatrics*. 2014;59(2):295-9.
28. Landi F, Liperoti R, Russo A, Giovannini S, Tosato M, Capoluongo E, et al. Sarcopenia as a risk factor for falls in elderly individuals: results from the iSIRENTE study. *Clinical nutrition (Edinburgh, Scotland)*. 2012;31(5):652-8.
29. Scott D, Hayes A, Sanders KM, Aitken D, Ebeling PR, Jones G. Operational definitions of sarcopenia and their associations with 5-year changes in falls risk in community-dwelling middle-aged and older adults. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2014;25(1):187-93.
30. 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. *Quality of Life Research*. 2018;27(5):1171-9.
31. 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. *Journal of clinical epidemiology*. 2010;63(7):737-45.
32. Baumgartner TA. Norm-referenced measurement: reliability. *Measurement concepts in physical education and exercise science*. 1989;20:45-7.
33. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology)*: Cambridge: Cambridge University Press.; 2011.
34. Overend T, Anderson C, Sawant A, Perryman B, Locking-Cusolito H. Relative and absolute reliability of physical function measures in people with end-stage renal disease. *Physiotherapy Canada Physiotherapie Canada*. 2010;62(2):122-8.
35. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *Journal of clinical epidemiology*. 2006;59(10):1033-9.
36. Liaw LJ, Hsieh CL, Lo SK, Chen HM, Lee S, Lin JH. The relative and absolute reliability of two balance performance measures in chronic stroke patients. *Disabil Rehabil*. 2008;30(9):656-61.
37. Fletcher RH, Fletcher SW, Fletcher GS. *Clinical epidemiology : the essentials*. Fifth edition: Lippincott Williams & Wilkins, a Wolters Kluwer business; 2015.
38. Hopewell S, Copsey B, Nicolson P, Adedire B, Boniface G, Lamb S. Multifactorial interventions for preventing falls in older people living in the community: a systematic review and meta-analysis of 41 trials and almost 20 000 participants. *Br J Sports Med*. 2019.

39. Ploutz-Snyder LL, Manini T, Ploutz-Snyder RJ, Wolf DA. Functionally relevant thresholds of quadriceps femoris strength. *The journals of gerontology Series A, Biological sciences and medical sciences*. 2002;57(4):B144-52.
40. Hasselgren L, Olsson LL, Nyberg L. Is leg muscle strength correlated with functional balance and mobility among inpatients in geriatric rehabilitation? *Archives of gerontology and geriatrics*. 2011;52(3):e220-5.
41. Marko M, Neville CG, Prince MA, Ploutz-Snyder LL. Lower-extremity force decrements identify early mobility decline among community-dwelling older adults. *Physical therapy*. 2012;92(9):1148-59.
42. Palmer TB, Thiele RM, Williams KB, Adams BM, Akehi K, Smith DB, et al. The identification of fall history using maximal and rapid isometric torque characteristics of the hip extensors in healthy, recreationally active elderly females: a preliminary investigation. *Aging clinical and experimental research*. 2015;27(4):431-8.
43. Morcelli MH, Rossi DM, Karuka AH, Crozara LF, Hallal CZ, Marques NR, et al. Peak torque, reaction time, and rate of torque development of hip abductors and adductors of older women. *Physiotherapy theory and practice*. 2016;32(1):45-52.
44. Morcelli MH, Crozara LF, Rossi DM, LaRoche DP, Marques NR, Hallal CZ, et al. Hip muscles strength and activation in older fallers and non-fallers. *Isokinetics and Exercise Science*. 2014;22:191-6.
45. Arnold CM, Warkentin KD, Chilibeck PD, Magnus CR. The reliability and validity of handheld dynamometry for the measurement of lower-extremity muscle strength in older adults. *J Strength Cond Res*. 2010;24(3):815-24.
46. Johnson ME, Mille ML, Martinez KM, Crombie G, Rogers MW. Age-related changes in hip abductor and adductor joint torques. *Archives of physical medicine and rehabilitation*. 2004;85(4):593-7.
47. Batista FS, Gomes GA, D'Elboux MJ, Cintra FA, Neri AL, Guariento ME, et al. Relationship between lower-limb muscle strength and functional independence among elderly people according to frailty criteria: a cross-sectional study. *Sao Paulo medical journal = Revista paulista de medicina*. 2014;132(5):282-9.
48. Hilliard MJ, Martinez KM, Janssen I, Edwards B, Mille ML, Zhang Y, et al. Lateral balance factors predict future falls in community-living older adults. *Archives of physical medicine and rehabilitation*. 2008;89(9):1708-13.
49. Arvin M, van Dieen JH, Faber GS, Pijnappels M, Hoozemans MJ, Verschueren SM. Hip abductor neuromuscular capacity: A limiting factor in mediolateral balance control in older adults? *Clinical biomechanics (Bristol, Avon)*. 2016;37:27-33.
50. Arvin M, Hoozemans MJ, Burger BJ, Rispens SM, Verschueren SM, van Dieen JH, et al. Effects of hip abductor muscle fatigue on gait control and hip position sense in healthy older adults. *Gait & posture*. 2015;42(4):545-9.
51. Cofre Lizama LE, Pijnappels M, Faber GH, Reeves PN, Verschueren SM, van Dieen JH. Age effects on mediolateral balance control. *PloS one*. 2014;9(10):e110757.
52. Nankaku M, Kanzaki H, Tsuboyama T, Nakamura T. Evaluation of hip fracture risk in relation to fall direction. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2005;16(11):1315-20.
53. Hayes WC, Myers ER, Morris JN, Gerhart TN, Yett HS, Lipsitz LA. Impact near the hip dominates fracture risk in elderly nursing home residents who fall. *Calcified tissue international*. 1993;52(3):192-8.
54. Greenspan SL, Myers ER, Kiel DP, Parker RA, Hayes WC, Resnick NM. Fall direction, bone mineral density, and function: risk factors for hip fracture in frail nursing home elderly. *The American journal of medicine*. 1998;104(6):539-45.
55. Daun F, Kibele A. Different strength declines in leg primary movers versus stabilizers across age-Implications for the risk of falls in older adults? *PloS one*. 2019;14(3):e0213361.
56. Choi WJ, Crompton PA, Robinovitch SN. Effects of hip abductor muscle forces and knee boundary conditions on femoral neck stresses during simulated falls. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2015;26(1):291-301.

CHAPTER 2

Evaluation of hip abductor and adductor strength in the elderly: a reliability study

Simone C. Gafner, Caroline H. G. Bastiaenen, Philippe Terrier, Ilona Punt, Serge Ferrari, Gabriel Gold, Rob de Bie, Lara Allet

Eur Rev Aging Phys Act Off J Eur Group Res Elder Phys Act. 2017;14:5.
DOI: 10.1186/s11556-017-0174-6

ABSTRACT

Background:

In elderly individuals, an increased muscle strength contributes to the diminution of the falls risk and associated adverse events. An increasing interest in lateral control exists due to the fatal consequences of postero-lateral falls. Therefore, a proper assessment of frontal plane hip muscle strength in elderly is important but remains challenging. Therefore, we aimed to investigate the feasibility and repeatability of a hip abductor and adductor maximum voluntary isometric strength (MVIS) and rate of force generation (RFG) test in elderly. This represents an initial step in the development process of a new and clinically relevant test that could lead to more specific treatment protocols for this population.

Methods:

In this measurement focused study hip abduction (ABD) and adduction (ADD) MVIS and RFG were tested twice within one to three hours with a dynamometer fixed to a custom made frame in a geriatric population including fallers and non-fallers. Intraclass correlation coefficient ($ICC_{\text{agreement}}$), standard error of measurement (SEM), and smallest detectable difference (SDD) were determined.

Results:

All recruited persons ($N = 76$; mean age (SD) 80.46 (7.05) years old) completed the tests. The average time needed to complete the strength tests was 10.58 min. (1.56) per muscle group. The reliability of the hip ABD and ADD was high with $ICC_{\text{agreement}}$ ranging from 0.83 to 0.97. The SDD varied between 18.1 and 81.8% depending on the muscle group and type of strength that was evaluated.

Conclusion:

Hip abductor and adductor strength measures in older person are feasible and reliable. However, the significance of moderate changes in these measurements may be limited by the large SDD and SEM. Therefore, physical therapist should be careful when using this measure for assessing the progress of an individual person in a daily clinical use.

Keywords: Geriatrics, Elderly, Muscle strength, Hip, Test, Measurements, Reliability

BACKGROUND

Approximately 50% of Swiss people older than 75 years live alone and most of them show a clinically relevant decline in their ability to execute daily life activities (1). These disabilities are linked to decreased physiological capabilities, such as diminished muscle strength and power production. Reduction in muscle strength in the elderly increases the risk of mobility limitation (2–8), falls (9) and mortality (10–13).

Shumway-Cook et al. (14) showed that the annual health care costs were \$2,000 (29%) higher in older adults reporting one fall during the previous year comparing to non-fallers and \$5,600 (79%) higher among those reporting recurrent falls during the previous year. Therefore, the maintenance of muscle strength and power with advancing age is of great clinical importance as it contributes to diminish the risk of falls, fractures and associated adverse events, and hence, is implicated when trying to maintain independence (15, 16).

The role of lower limb muscle groups during walking patterns has been described by Perry (17). Frontal plane hip muscles (abductors/adductors) play an important role during walking as these muscles are essential to control the head, arms and trunk during single leg stance phase. They further allow a proper swing foot placement after the swing phase (17, 18). An increasing interest in lateral control is emerging, given that impairments in this domain are thought to increase falls risk (19, 20). Recent findings in nonlinear gait analysis even suggest that gait stability in the frontal plane is of interest for predicting fall risk (21, 22). Finally, lateral and postero-lateral falls have greater hip injury potential, including hip fractures, than falls in other directions (23–25).

Our previous work showed that frontal plane hip strength might be able to compensate for distal neuromuscular deficits in persons with distal symmetric neuropathy (DSP) during gait in challenging circumstances as it did during unipedal stance time (UST) (26). Muscle strength (maximum voluntary isometric strength (MVIS)) was particularly important for static tasks like balancing on one leg, whereas the power (rate of force generation (RFG)) was important for safe ambulation (19, 26, 27). The absence of therapy to restore nerve health in people with age-related declining peripheral nerve function makes it important to find innovative strategies, which allow patients to compensate for these distal nerve function losses. The preservation of the ability to increase hip strength in these patients might enable them to compensate these distal sensory deficits through other bodily functions and hence, safely maintain mobility and prevent falls.

However, effective and correct assessment of frontal plane hip muscle strength and power in the elderly remains a challenge. The clinical experience show that an assessment with an isokinetic system like the Biodex is time consuming and

it can be difficult to install older people on such isokinetic apparatus. In addition, only few rehabilitation centers have such equipment. Therefore daily clinical use of the biodex apparatus to measure hip strength in older persons, in particular hip abduction and adduction strength which necessitates a side-lying or supine position, seems unrealistic. The fact that only a few studies exist (28, 29) that measure hip abduction and adduction strength in older persons at risk of falling underlines the difficulty of properly measuring hip frontal plane strength. The aim of this study was to investigate the feasibility and repeatability of a newly developed hip abductor and adductor MVIS and RFG test with a dynamometer fixed to a custom made frame in a geriatric population including fallers and non-fallers. This is an important first step in the development process of this innovative measure for clinical practice as it could lead to the development of more specific treatment protocols for this population. To do so we recruited men and women ≥ 65 years of age, both with and without a history of falls, and assessed: 1) how many of the recruited persons could successfully finish the hip abductor and adductor MVIS and RFG test, 2) the time needed to administer the test, 3) the test-retest reliability, 4) the measurement error (SEM) as well as 5) the smallest detectable difference (SDD).

We hypothesized that the hip abductor and adductor MVIS and RFG test with a dynamometer fixed to a custom made frame is feasible and would reliably assess hip strength in a population consisting of both fallers and non-fallers with intraclass correlation coefficients (ICC's_{agreement}) of ≥ 0.7 .

METHODS

Subjects and recruitment

Overall, 76 older persons were recruited from a geriatric hospital that is part of the Geneva University Hospitals and in an outpatients practice in Switzerland. Half of them (38 persons consisting in 19 fallers and 19 non-fallers) underwent the abduction and the other half of them (38 persons consisting in 19 fallers and 19 non-fallers) the adduction test.

To be included participants had to be over 65 years. They were excluded if their medical record contained a history or evidence of any significant central nervous system dysfunction (i.e., hemiparesis, myelopathy or cerebellar ataxia), any neuromuscular disorder other than a distal symmetric peripheral neuropathy (i.e., myopathy or myasthenia gravis) or evidence of vestibular dysfunction. In addition, we excluded persons with a moderate or severe dementia (Minimal Mental State Exam (MMSE) < 18) which did not allow them to understand the study information and the necessary instructions, and persons with a severe sepsis, metastatic cancer,

angina or angina-equivalent symptoms with exercise. We also excluded persons with a plantar skin sore or joint replacement within the previous year and persons with a non-consolidated fracture, significant musculoskeletal deformity (i.e., amputation, Charcot changes), lower limb or spinal arthritis or pain that limited proper execution of the test. Within the recruited population a faller was defined as a person who had one or more falls during the last 12 months. Non-fallers were defined as participants who did not fall during the last 12 months. A fall was defined as an event, which results in a person coming to rest inadvertently on the ground or floor or other lower level (30).

Ethics

The study was approved by the ethical commission in Geneva (CCER - 14-235). All participants signed the written informed consent (declaration of Helsinki) after having received information about the study and time to decide about participation.

Dynamometer

An analog dynamometer (SENSIX®, Poitiers, France) able to measure forces between 0 to 667 N with a precision of 0.002 N was used to measure hip abductor strength (N). The calibrated analog dynamometer was coupled with the DELSYS System (Trigno sensor, DELSYS, INC Boston; MA) that digitalized the analog output (3.3 V) with a sampling rate of 1926 Hz and a 16 bit resolution.

Procedure

The force of hip abductors and adductors was measured in side-lying position by an experienced physical therapist. This position was previously described as the most valid and reliable position to measure hip abductor strength in young people (29). In addition, the influence of hip abductor strength of the contra-lateral leg is reduced in the side-lying position compared to the supine or standing position (29). The same examiner repeated the whole test procedure on the same day with a break of minimum one and maximum three hours in between. This time was estimated as being enough to fully recover, but not too long to have a change in the strength performance of the elderly frail persons.

Hip Abduction Test

For the hip abduction test, the participant was in a side-lying position. The tester held the participant's leg in a five to eight degree abduction position, close to the

dynamometer, which was attached to a frame fixed on the bed. The dynamometer was positioned in a 10° angled position to the horizontal line, five cm proximal to the malleolus externus (31). The participant was instructed to push his leg as quickly and forcefully as possible towards the dynamometer, to hold it with his maximum force for three seconds and to relax (Figure 1a). Verbal encouragement was given during all measurements.

Figure 1 a) For the hip abduction test the participant was in a side-lying position. The dynamometer was positioned in a 10° angled position to the horizontal line, five cm proximal to the malleolus externus. **b)** For hip adduction test the person was again in a side-lying position. This time the lower leg of the participant was tested. The starting position is at 0° adduction.

Hip Adduction Test

For hip adduction testing, the person was again in a side-lying position. This time the lower leg of the participant was tested. The starting position was at 0° adduction and the participant was asked to push the leg in five degree adduction against the dynamometer fixed to a frame (Figure 1b).The participant had to repeat each test three times with a break of one minute between every trial. Verbal encouragement was given during all measurements.

Data processing

The raw force signals were exported to Matlab (Mathworks, Natick, MA), which was used for data analysis and statistics. The signal was low-pass filtered (75 ms moving average) to attenuate high-frequency noise. The MVIS was defined as the peak value reached within zero to four seconds. The rate of force change was evaluated over 50 ms after 10% of the MVIS was reached (see Figure 2 for graphical explanations). Both MVIS and RFG were normalized by body mass (32, 33).

Figure 2: The rate of force change was evaluated over 50 ms after 10% of the MVIS was reached

Statistics

We used descriptive statistics to summarize the characteristics of the study population. The number of participants who could successfully finish the hip abductor and adductor MVIS and RFG test are expressed as a percentage and the time to administer the test is expressed in mean and SD. Scatter plots were used to display the individual results of both testing sessions. The identity line (test = retest) is shown to get a better estimation of the dispersion of the results and of potential

bias between sessions. ICC_{agreements} were used to assess the repeatability of each dependent variable (MVIS, RFG) between both measurement sessions. ICC_{agreement} represents the proportion of between-subject variance as compared to total variance. Three ICC's_{agreement} were computed in Abduction and Adduction groups: one from the whole sample (total), one for the subgroup of fallers, and one for the subgroup of non-fallers. The averages of the six trials within each session (three trials with the left leg and three trials with the right leg) were used as two within- subject repetitions. The ICC method was that proposed by McGraw and Wong (34). We applied the ICC(A,1) model, which assesses the degree of agreement among measurements assuming a two-way random effects model. In addition, we computed 95% confidence intervals (CI) through bootstrapping (5000 resamples, bias corrected and accelerated percentile method). In addition the group level estimation of the within-subject average variability (35), i.e., the standard error of measurements (SEM), was computed as follows: $SEM=S_r \sqrt{1-ICC}$, where S_r is the standard deviation of the whole sample including both sessions. The smallest detectable difference (SDD) was also computed: $SDD = SEM \times 1.96 \times \sqrt{2}$, it is the smallest change that could be considered significant. The SDD was normalized by the mean and expressed as percentage.

RESULTS

Overall 76 older subjects were recruited. Half of them (19 fallers and 19 non-fallers) underwent the abduction and the other half (19 fallers and 19 non-fallers) the adduction test. The characteristics of the whole study population as well as the characteristics per subgroup are presented in Table 1.

All recruited persons could successfully complete the tests. To measure one muscle group (MVIS and RFG hip abduction or adduction of both sides (left and right)) a mean time of 10.58 ± 1.56 min was necessary. The reliability of the hip abduction and adduction test is represented in Table 2 and Figure 3. The ICC's_{agreement} for hip abduction were 0.84 [0.72–0.94] for fallers and 0.97 [0.94–0.99] for non-fallers for the MVIS and 0.85 [0.70–0.93] for fallers and 0.96 [0.92–0.98] for non- fallers for the RFG measures. The ICC's_{agreement} for hip adduction were 0.85 [0.70–0.93] for fallers and 0.96 [0.92–0.98] for non-fallers for the MVIS and 0.83 [0.72– 0.94] for fallers and 0.91 [0.76–0.97] for non-fallers for the RFG measures.

The SEM and SDD for hip abduction MVIS measures of the total sample was 0.12 N/kg, respectively 32.6% and 1.05 N/kg/s, respectively 51.7% for hip abduction RFG measures. The SEM and SDD for hip adduction MVIS measures of the total sample was 0.17 N/kg, respectively 39.6% and 0.77 N/kg/s, respectively 48.9% for hip adduction RFG measures.

Table 1: Description in mean (SD) of the whole population and the two subgroups recruited for the hip abduction and hip adduction strength test

Study Population	All (N = 76)	Non-fallers (N = 38)	Fallers (N = 38)
Sex: F/M	41/35	22/16	19/19
Age (years)	80.46 (7.05)	78.47 (6.64)	82.45 (6.98)
Weight (kg)	67.58 (12.58)	69.25 (10.68)	65.92 (14.18)
BMI	24.89 (3.95)	25.00 (3.04)	24.78 (4.73)
SPPB ^a	8.45 (3.07)	10.45 (2.37)	6.41 (2.27)
Subgroup for abduction test	Non-fallers (N = 19)	Fallers (N = 19)	P-value
Sex: F/M	11/8	9/10	
Age (years)	79.84 (7.08)	81.37 (7.67)	0.528
Weight (kg)	68.15 (12.01)	66.17 (15.15)	0.658
BMI	24.71 (3.39)	23.84 (4.40)	0.500
SPPB	10.00 (2.54)	6.37 (2.03)	<0.001
ABD MVIS (N/kg)	1.23 (0.49)	0.77 (0.37)	<0.01
ABD RFG (N/kg/s)	7.65 (5.27)	3.498 (2.51)	<0.01
Subgroup for adduction test	Non-Fallers (N = 19)	Fallers (N = 19)	P-value
Sex: F/M	11/8	10/9	
Age (years)	77.11 (6.04)	83.53 (6.23)	<0.01
Weight (kg)	70.35 (9.36)	65.66 (13.55)	0.222
BMI	25.28 (2.72)	25.72 (4.99)	0.743
SPPB ^a	10.89 (2.16)	6.44 (2.55)	<0.001
ADD MVIS (N/kg)	1.40 (0.43)	0.93 (0.51)	<0.01
ADD RFG (N/kg/s)	5.61 (2.99)	2.80 (2.46)	<0.01

Abd Abduction, *Add* Adduction, *MVIS* Maximum Voluntary Isometric Strength, *RFG* Rate of Force Generation, *N* Newton, *kg* Kilogram, *s* Seconds, *SPPB* Short Physical Performance Battery

^aOne faller of the adduction group didn't realize this test

Subgroup analyses showed a SEM for hip abduction MVIS measures of 0.15 N/kg for fallers and 0.08 N/kg for non-fallers and SEM for hip abduction RFG measures of 1.12 N/kg/s for fallers and 1.01 N/kg/s for non-fallers. The SDD of hip abduction MVIS measures was 51.8% for fallers and 18.1% for non-fallers. The SDD of hip abduction RFG measures was 81.8% for fallers and 37.2% for non-fallers.

For the subgroup who performed the hip adduction test the SEM of MVIS measures was 0.16 N/kg for fallers and of 0.19 N/kg for non-fallers; respectively of 0.72 N/kg/s for fallers and of 0.83 N/kg/s for non-fallers for RFG measures. The SDD was 44.9% for fallers and 39.8% for non-fallers for MVIS measures and 67.7%, respectively 45.9% for RFG measures.

The scatterplot (Figure 3a and b) shows the individual results of both testing sessions and gives an estimation of the dispersion of the results. We observed that more points are above the identity line than below.

Table 2: Intraclass correlation coefficients (ICC's agreement), Standard error of measurement (SEM) and smallest detectable change (SDD) of the hip abductor and hip adductor strength test

	Fallers (N = 19)			Non-Fallers (N = 19)		
	ICC [95%CI]	SEM [unit of measure]	SDD [%]	ICC [95%CI]	SEM [unit of measure]	SDD [%]
ABD						
MVIS	0.94 [0.87–0.97]	0.12	32%	0.84 [0.72–0.94]	0.15	51%
RFG	0.94 [0.91–0.97]	1.05	51%	0.85 [0.70–0.93]	1.12	81%
ADD						
MVIS	0.90 [0.84–0.95]	0.17	39%	0.90 [0.72–0.96]	0.16	44%
RFG	0.94 [0.90–0.97]	0.77	48%	0.93 [0.84–0.97]	0.72	67%

ABD = Abduction; ADD = Adduction, MVIS = Maximum Voluntary Isometric Strength (N/kg); RFG= Rate of Force Generation (N/kg/s)

Figure. 3 a and b: Scatterplot indicating the relationship between the first and second test of a) hip abduction strength respectively, b) adduction strength for fallers and non-fallers

DISCUSSION

The results of this study show that the measure of hip abductor and adductor strength in elderly people is practicable with a comparable good reliability in fallers and non-fallers. The ICC's_{agreement} can be explained by a high inter- individual variability in hip abduction and adduction strength in our population which minimizes the influence of the intra-individual variability in the population and hence, allows to recommend this measurement approach for research in a comparable population. The set-up that we used to measure hip abductor and adductor strength was the same as the set-up used by Widler et al. and Nadler et al. (28, 29). Widler et al. (29) tested the reliability of unilateral hip abductor strength assessments in sixteen healthy young subjects in three different body positions with the use of a stabilized commercial dynamometer in the side-lying, supine, and standing positions. The highest MVIS value for each side and position was retained and strength data provided by the dynamometer

were consistently normalized to body weight. The retest took place 48 to 72h after the first assessment and it was assumed that the participants' characteristics did not change in that time interval. The maximal hip abductor strength was significantly higher in the side-lying position compared with the standing and supine positions. The test-retest reliability of strength measurements in terms of $ICC_{\text{agreement}}$ in their young population was of 0.90 in the side-lying position which is very similar to the $ICC_{\text{agreement}}$ that we calculated in our population ($ICC_{\text{agreement}} = 0.94$). However, 95% confidence intervals were not presented which would allow a full comparison with our population. Nadler et al. (28) also assessed the reliability of hip abductor strength in ten subjects aged from 25 to 35 years. In their study an inexperienced examiner did three consecutive measures. The procedure was repeated two weeks later by the same subjects with an evaluator blinded to the initial results. For each of the hip muscles tested, an average and maximal value was computed and presented from the three test repetitions obtained. The $ICC_{\text{agreement}}$ in their study were again similar to the results that we had in older adults and ranged from 0.95 to 0.98. Again, these authors did not present a 95% confidence interval of their $ICC_{\text{agreement}}$ which would allow a comparison of the reliability obtained for their young and our older adults. In our study we computed the mean value of three left and three right trials before calculating the $ICC_{\text{agreement}}$. Additionally, we also calculated the $ICC_{\text{agreement}}$ based on the trial with the best performance and compared it with the $ICC_{\text{agreement}}$ calculated based on the average. No relevant differences were found between these two approaches. Widler et al. and Nadler et al. (28, 29) did not assess the reliability of the hip abduction RFG and the hip adduction MVIS and RFG, therefore, we could not compare our values with theirs. Nevertheless, reliability outcomes of hip adductor MVIS and reliability coefficients of RFG measures in our study were similar to reliability outcomes of the hip abduction MVIS. In addition, Widler et al. and Nadler et al. (28, 29) did not present the SDD or the SEM which could give an indication if the test can be used in daily practice. The high inter-subject variability for hip abduction strength measures in our population and the relatively high inter-subject variability in hip abduction strength measures in healthy young subjects confirm that hip abductor strength is a variable with clinical relevance. Nadler et al. (28) calculated a coefficient of variation for hip abduction strength of 26% in young subjects whereas the inter-subject variability in our study with older persons was of 49%. The SEM, i.e., the intra-subject variability, was high in our study. For the overall group we had a SEM of 0.12 N/kg, respectively 1.05 N/kg/s for the hip abduction and of 0.19 N/kg, respectively 0.77 N/kg/s for the hip adduction. The SDD for hip abduction in the overall group was 32.6% foVIS and 51.6% for the RFG. For adduction, the values were 43.2 and 56.1% respectively. The variability that we also observed between the three measures of one single test session (results not shown) confirms that the

performance of older adults varies a lot. In addition, we could observe a small (not significant) learning effect between the first and second test series which might have contributed to the moderate to large SEM and SDD.

Even if the $ICC_{\text{agreement}}$ were high, physical therapists should be careful when using this measure for assessing the progress of an individual person in a daily clinical use due to the high SEMs and SDDs (35). The progress of an individual person in hip abduction strength must be above 32% for the MVIS and above 39% for the RFG to be considered as real. Still, the clinical relevance of changes greater than the SDD remain to be evaluated.

One solution could be to combine several tests to increase reliability. Through the Spearmann-Brown prophecy formula, we can predict that averaging two separated measurement sessions would improve SDD for adduction from 39 to 29% for the MVIS and from 48 to 35% for the RFG (36).

The strength of this study is that we used a test procedure which is feasible and which can be tolerated by older persons. No specific training is necessary. Compared to the isokinetic strength assessment tools, our procedure is less time-consuming and more practical for a geriatric population, and equally reliable to isokinetic strength assessments in young soccer player (37). In addition, we provide all necessary statistics to get a clear overview of the reliability of the test procedure.

The following limitations should be noted. We only assessed the intra-tester reliability and performed the two tests on the same day. The assessor was not blinded to the subgrouping (faller/non-faller). The fact that we observed more points above the identity line than below for hip adduction might indicate a small learning effect. In addition, the current set-up and the high SEM and SDD do not indicate that the test is easy to use in a daily setting. Even if the time needed to test a muscle group is short (approximately 10 min), this remains slightly too long for daily clinical use. Due to this, we should rethink the testing position, as the installation of the person in a side-lying position takes at least half of the time needed to perform the whole test. In addition, it would be pertinent to assess how much a patient can progress in his MVIS and RFG in order to make sure that our measure can detect this change. It might be that more practice trials would decrease the SEM and SDD and thus improve the sensitivity to measure a change.

Future studies should be done to investigate different elements of reliability in a larger sample to repeat our findings. In addition, validity studies on different aspects should be done. Validity studies should particularly focus on discriminant validity in order to assess if hip frontal strength measures are able to discriminate between elderly fallers and non-fallers. The differences observed in this study between these subgroups would suggest that hip frontal plane strength is of clinical relevance. If this can be confirmed such a test would enable clinicians to gather crucial information

about a directly targetable and trainable parameter related to fall risk. In addition, researchers should investigate the ability of hip abductor and adductor strength measures to assess changes over time after a specific strength training program (treatment effect). We are aware that discrimination and evaluation of change are two totally different abilities and that it is difficult for a measure to combine both features in an optimal way. Nevertheless, both abilities are relevant in the context of fall prevention and should, therefore, not be neglected.

CONCLUSION

In conclusion, we can say that measures of hip abductor and adductor strength in older persons are feasible and reliable. We encourage physical therapists to routinely assess hip frontal plane strength as it provides an interesting new goal for clinical practice, which could lead to the development of more specific treatment protocols for this population. However, the significance of moderate changes in these measurements may be limited by the large SDD and SEM. Therefore, physical therapists should be careful when using this measure for assessing the progress of an individual person in a daily clinical.

ABBREVIATIONS

ABD: Abduction; ADD: Adduction; DSP: Distal Symmetric Neuropathy; ICC: Intraclass Correlation Coefficient; MVIS: Maximum Voluntary Isometric Strength; RFG: Rate of Force Generation; SDD: Smallest Detectable Difference; SEM: Standard Error of Measurement; SPPB: Short Physical Performance; UST: Uni-pedal Stance Time

ACKNOWLEDGEMENTS

We would like to acknowledge all physical therapists of the geriatric hospital in Geneva and of the outpatients practice who helped us to recruit older person who fulfilled the inclusion criteria for this study.

FUNDING

The study was funded from an internal grant of the University Of Applied Sciences of Western Switzerland.

REFERENCES

- Schellhorn M, Stuck AE, Minder CE, Beck JC. Health Services Utilization of Elderly Swiss: Evidence from Panel Data. *Econ Health*. 2000;9(6):533–45.
- Manini TM, Visser M, Won-Park S, Patel KV, Strotmeyer ES, Chen H, Goodpaster B, De Rekeneire N, Newman AB, Simonsick EM, et al. Knee extension strength cutpoints for maintaining mobility. *J Am Geriatr Soc*. 2007;55(3):451–7.
- Visser M, Deeg DJ, Lips P, Harris TB, Bouter LM. Skeletal muscle mass and muscle strength in relation to lower-extremity performance in older men and women. *J Am Geriatr Soc*. 2000;48(4):381–6.
- Visser M, Harris TB, Fox KM, Hawkes W, Hebel JR, Yahiro JY, Michael R, Zimmerman SI, Magaziner J. Change in muscle mass and muscle strength after a hip fracture: relationship to mobility recovery. *J Gerontol A Biol Sci Med Sci*. 2000;55(8):M434–40.
- Visser M, Goodpaster BH, Kritchevsky SB, Newman AB, Nevitt M, Rubin SM, Simonsick EM, Harris TB. Muscle mass, muscle strength, and muscle fat infiltration as predictors of incident mobility limitations in well-functioning older persons. *J Gerontol A Biol Sci Med Sci*. 2005;60(3):324–33.
- Ferrucci L, Penninx BW, Volpato S, Harris TB, Bandeen-Roche K, Balfour J, Leveille SG, Fried LP, Md JM. Change in muscle strength explains accelerated decline of physical function in older women with high interleukin-6 serum levels. *J Am Geriatr Soc*. 2002;50(12):1947–54.
- Hasselgren L, Olsson LL, Nyberg L. Is leg muscle strength correlated with functional balance and mobility among inpatients in geriatric rehabilitation? *Arch Gerontol Geriatr*. 2011;52(3):e220–5.
- Ploutz-Snyder LL, Manini T, Ploutz-Snyder RJ, Wolf DA. Functionally relevant thresholds of quadriceps femoris strength. *J Gerontol A Biol Sci Med Sci*. 2002;57(4):B144–52.
- Moreland JD, Richardson JA, Goldsmith CH, Clase CM. Muscle weakness and falls in older adults: a systematic review and meta-analysis. *J Am Geriatr Soc*. 2004;52(7):1121–9.
- Newman AB, Kupelian V, Visser M, Simonsick EM, Goodpaster BH, Kritchevsky SB, Tylavsky FA, Rubin SM, Harris TB. Strength, but not muscle mass, is associated with mortality in the health, aging and body composition study cohort. *J Gerontol A Biol Sci Med Sci*. 2006;61(1):72–7.
- Xue QL, Beamer BA, Chaves PH, Guralnik JM, Fried LP. Heterogeneity in rate of decline in grip, hip, and knee strength and the risk of all-cause mortality: the Women's Health and Aging Study II. *J Am Geriatr Soc*. 2010;58(11):2076–84.
- Takata Y, Ansai T, Soh I, Awano S, Yoshitake Y, Kimura Y, Nakamichi I, Goto K, Fujisawa R, Sonoki K, et al. Physical fitness and 6.5-year mortality in an 85-year-old community-dwelling population. *Arch Gerontol Geriatr*. 2012;54(1):28–33.

13. Artero EG, Lee DC, Ruiz JR, Sui X, Ortega FB, Church TS, Lavie CJ, Castillo MJ, Blair SN. A prospective study of muscular strength and all-cause mortality in men with hypertension. *J Am Coll Cardiol*. 2011;57(18):1831–7.
14. Shumway-Cook A, Ciol MA, Hoffman J, Dudgeon BJ, Yorkston K, Chan L. Falls in the Medicare population: incidence, associated factors, and impact on health care. *Phys Ther*. 2009;89(4):324–32.
15. British Geriatrics Society aAAoOSPofP. Guideline for the prevention of falls in older persons. American Geriatrics Society. *J Am Geriatr Soc*. 2001;49(5):664–72.
16. Kenny RA, Rubenstein LZ, Tinetti ME, Brewer K, Cameron KA, Capezuti EA, John DP, Lamb S, Martin F, Rockey PH, Suther M, Peterson E, Susskind O, Radcliff S, Addleman K, Drootin M, Ickowicz E, Lundebjerg N. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. *J Am Geriatr Soc*. 2011;59(1):148–57. doi:10.1111/j.1532-5415.2010.03234.x.
17. Perry J, editor. *Gait Analysis: Normal and pathological function*. 1st ed. 1992.
18. Vellas BJ, Rubenstein LZ, Ousset PJ, Faisant C, Kostek V, Nourhashemi F, Allard M, Albarede JL. One-leg standing balance and functional status in a population of 512 community-living elderly persons. *Aging (Milano)*. 1997;9(1–2):95–8.
19. Allet L, Kim H, Ashton-Miller J, De Mott T, Richardson JK. Step length after discrete perturbation predicts accidental falls and fall-related injury in elderly people with a range of peripheral neuropathy. *J Diabetes Complications*. 2014;28(1):79–84.
20. Richardson JK, Demott T, Allet L, Kim H, Ashton-Miller JA. Hip strength: ankle proprioceptive threshold ratio predicts falls and injury in diabetic neuropathy. *Muscle Nerve*. 2014;50(3):437–42.
21. Reynard F, Vuadens P, Deriaz O, Terrier P. Could local dynamic stability serve as an early predictor of falls in patients with moderate neurological gait disorders? A reliability and comparison study in healthy individuals and in patients with paresis of the lower extremities. *PLoS One*. 2014;9(6):e100550.
22. Terrier P, Reynard F. Effect of age on the variability and stability of gait: a cross-sectional treadmill study in healthy individuals between 20 and 69 years of age. *Gait Posture*. 2015;41(1):170–4.
23. Nankaku M, Kanzaki H, Tsuboyama T, Nakamura T. Evaluation of hip fracture risk in relation to fall direction. *Osteoporos Int*. 2005;16(11):1315–20.
24. Hayes WC, Myers ER, Morris JN, Gerhart TN, Yett HS, Lipsitz LA. Impact near the hip dominates fracture risk in elderly nursing home residents who fall. *Calcif Tissue Int*. 1993;52(3):192–8.
25. Greenspan SL, Myers ER, Kiel DP, Parker RA, Hayes WC, Resnick NM. Fall direction, bone mineral density, and function: risk factors for hip fracture in frail nursing home elderly. *Am J Med*. 1998;104(6):539–45.
26. Allet L, Kim H, Ashton-Miller J, De Mott T, Richardson JK. Frontal plane hip and ankle sensorimotor function, not age, predicts unipedal stance time. *Muscle Nerve*. 2012;45(4):578–85.
27. Allet L, Kim H, Ashton-Miller JA, Richardson JK. Which lower limb frontal plane sensory and motor functions predict gait speed and efficiency on uneven surfaces in older persons with diabetic neuropathy? *PM R*. 2012; 4(10):726–33.
28. Nadler SF, DePrince ML, Hauesien N, Malanga GA, Stitik TP, Price E. Portable dynamometer anchoring station for measuring strength of the hip extensors and abductors. *Arch Phys Med Rehabil*. 2000;81(8):1072–6.
29. Widler KS, Glatthorn JF, Bizzini M, Impellizzeri FM, Munzinger U, Leunig M, Maffiuletti NA. Assessment of hip abductor muscle strength. A validity and reliability study. *J Bone Joint Surg Am*. 2009;91(11):2666–72.
30. WHO. World Health Organization report/WHO Global report on falls Prevention in older Age. 2007.
31. Bazett-Jones DM, Cobb SC, Joshi MN, Cashin SE, Earl JE. Normalizing hip muscle strength: establishing body-size-independent measurements. *Arch Phys Med Rehabil*. 2011;92(1):76–82.
32. Hsu AL, Tang PF, Jan MH. Test-retest reliability of isokinetic muscle strength of the lower extremities in patients with stroke. *Arch Phys Med Rehabil*. 2002;83(8):1130–7.
33. McCarthy EK, Horvat MA, Holtsberg PA, Wisenbaker JM. Repeated chair stands as a measure of lower limb strength in sexagenarian women. *J Gerontol A Biol Sci Med Sci*. 2004;59(11):1207–12.
34. McGraw KO, Wong S. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30.
35. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231–40.
36. Hollman JH, Childs KB, McNeil ML, Mueller AC, Quilter CM, Youdas JW. Number of strides required for reliable measurements of pace, rhythm and variability parameters of gait during normal and dual task walking in older individuals. *Gait Posture*. 2010;32(1):23–8.
37. Gerodimos V, Karatrantou K, Paschalis V, Zafeiridis A, Katsareli E, Bilios P, Kellis S. Reliability of concentric and eccentric strength of hip abductor and adductor muscles in young soccer players. *Biol Sport*. 2015;32(4):351–6.

CHAPTER 3

Intra-rater reliability of hip abductor isometric strength testing in a standing position in older fallers and non-fallers.

Anne-Violette Bruyneel, Simone C. Gafner, Serge Ferrari, Gabriel Gold, Dominique Monnin, Philippe Terrier, Caroline H. G. Bastiaenen, Lara Allet

European review of aging and physical activity: official journal of the European Group for Research into Elderly and Physical Activity. 2018;15:9.

DOI: 10.1186/s11556-018-0198-6

ABSTRACT

Background:

Reduced hip muscle strength has been shown to be a major factor related to falls in older persons. However, comprehensive assessment of hip abduction strength in the clinical setting is challenging.

Purpose:

The aim of this study was to investigate the feasibility and intra-rater reliability of a quick and simple hip abductor strength test in a functional standing position.

Methods:

Individuals over 65 years of age were recruited from the geriatric department of a university hospital and an outpatient clinic. Thirty-two older subjects, including 16 fallers (≥ 1 fall during the last 12 months) and 16 non-fallers were included. Maximum voluntary isometric strength (MVIS) and rate of force generation (RFG) of the hip abductors of the right leg were evaluated in a standing position using a hand-held dynamometer. Two test-sessions were carried out. All hip strength values were normalized to participants' weight. Reliability was determined using the intra-class correlation coefficient agreement ($ICC_{\text{agreement}}$), the standard error of measurement (SEM) and a Bland and Altman analysis (BA).

Results:

All participants completed the strength tests, which took a mean 2.47 ± 0.49 minutes (one limb). Intra-rater reliability was higher for MVIS (0.98 [95% CI: $0.95 - 0.99$]) than RFG ($ICC=0.93$ [$0.87 - 0.97$]) for the entire sample. In the non-fallers, ICC was 0.98 [$0.95 - 1.00$] (SEM= 0.08 N.kg^{-1}) for MVIS and 0.88 [$0.75 - 0.96$] for RFG (SEM= $1.34 \text{ N.kg}^{-1}.\text{s}^{-1}$). In the fallers, ICC was 0.94 [$0.89 - 0.98$] (SEM= 0.11 N.kg^{-1}) for MVIS and 0.93 [$0.84 - 0.98$] (SEM= $1.12 \text{ N.kg}^{-1}.\text{s}^{-1}$) for RFG. The BA plot showed that the MVIS and RFG values did not differ across test-sessions, showing that no learning effect occurred (no systematic effect). The mean differences between test-sessions were larger and the LOA smaller in the fallers than in the non-fallers.

Conclusion:

Assessment of hip strength in a standing position is feasible, rapid and reliable. We therefore recommend this position for clinical practice. Future studies should investigate the diagnostic value of hip abductor strength in standing to discriminate between fallers and non-fallers, and to determine if change in strength following a falls prevention program reduces the risk of falls.

Key words: older, falls, muscle strength, hip, abductor, standing, reliability

BACKGROUND

According to the World Health Organization, the proportion of people above the age of 60 years is growing more rapidly than any other age group (1). In 2025, around 1.2 billion people worldwide will be over the age of 60. Approximately 30% of people over 65 years of age fall each year. Falls lead to injuries, deconditioning, loss of independence and quality of life, and even death (2,3). Sixty-five percent of falls occur indoors, with 49% occurring while walking within the home (2).

The most recent Cochrane review showed that exercise programs that focus on improving balance and strength reduce the rate and risk of falls, but it is unclear which strength exercises and training modalities are the most effective (3). Weakness of the hip muscles has been shown to be a major factor related to falls in older persons (4). Maintaining hip muscle strength is thus important to decrease the risk of falls and associated fractures and adverse events. Comprehensive assessment of sensorimotor function is necessary to develop more effective treatments and improve understanding of the relationship between strength and falls. Hip abductor strength has been identified as a key parameter related to fall-risk (5,6). Arvin et al. showed that hip abductor muscles play an important role in medio-lateral balance control in older people (5). Proprioception of hip joint may be reduced if the hip abductors are fatigued (7). Hip muscle fatigue is associated with increased gait variability, step-to-step asymmetry in the frontal plane and slower medio-lateral trunk motion (7), all of which have been shown to be associated with an increased fall-risk (8–10). Addisson et al. also found an association between hip abductor strength and balance strategies in older adults (11). They showed that subjects with low hip abduction strength use a medial step strategy to recover balance from a perturbation in the frontal plane, whereas older adults with normal hip abductor strength use a cross-step strategy, which is more stable (11).

Given the impact of weak hip abductors on dynamic stability in older persons, a systematic, valid and reliable assessment of strength is necessary. However, the measurement of hip abductor muscle strength in the clinical setting is challenging and thus very few studies assessed psychometric properties of available strength tests (12–14). A previous study in our group highlighted the feasibility and good intra-rater reliability of a hip abductor strength test in a side-lying position in older people (13). In healthy young subjects, Widler et al. (12), showed that this position has the most valid and reliable results compared to supine and standing positions. Moreover, hip abductor strength measured in supine position had a better test-retest reliability in healthy older subjects compared to young subjects, especially when using a belt-resisted test (15). This higher reliability in older subjects could be attributed to a higher between-subjects strength variation than in younger persons. Moreover,

Kramer et al. suggested a higher risk of compensation during the test when the subject has strong abductor strength (15). So, the older would compensate less than the young, which would be favorable to better reliability. Thus, these two sources of variation seem to make the reliability outcome for older in that specific situation more favorable than for younger people. It is therefore likely that standing can provide a reliable test of hip abduction strength for this older population. Indeed, although measurements are often carried out in side-lying or supine, these positions are not functional as falls mostly occur while walking (2). In addition, the test procedure in side-lying is rather complicated for routine clinical use in older patients. Assessment in standing would be more functional and provide a better reflection of real life. In addition, this position is more practical in the clinical setting. Therefore, the challenge is to find a feasible and reliable hip abductor strength test in standing for older patients with balance disorders.

Most often clinical strength tests evaluate only the maximum strength value (12,15,16). In standing position, muscle strength (maximum voluntary isometric strength - MVIS) is particularly important for static tasks like keeping balance on one leg, whereas the rate of force generation (RFG) is important for safe ambulation (6,17,18). RFG is considered to be a parameter of the ability to rapidly generate strength and is an important component for joint stability and postural control (19). An increase of MVIS is not necessarily associated with an increase of RFG (20), hence the importance of using both parameters.

The purpose of this study was to investigate the feasibility and intra-rater reliability of a hip abductor strength test in a standing position using a hand-held dynamometer (HHD) in older people at risk of falls. More specifically we assessed: 1) how many participants completed the hip abductor test, 2) the time needed to complete the test, 3) test-retest reliability for the total group, and for the faller and non-faller sub-groups, 4) the standard error of measurement (SEM) as well as the smallest detectable difference (SDD), and 5) bias associated with limits of agreement analysis. When the analysis is performed for the whole group, the more heterogeneous strength values may increase the reliability. By adding the same analyses for both groups (fallers and non-fallers) with probably more homogeneous strength values that potential could lead to a decrease in outcomes (ICC) compared to the whole group we are able to receive a full picture of reliability in the whole group and in relation to both subgroups.

We hypothesized that the hip abductor strength test in standing would be feasible, rapid and have good intra-rater reliability in the whole group and the subgroups fallers and non-fallers separately ($ICC_{\text{agreement}} > 0.75$).

METHODS

Participants

Subjects aged over 65 years were consecutively recruited from the geriatric department of the Geneva University Hospital and an outpatient clinic in Switzerland. Thirty-two participants, including fallers and non-fallers, underwent the isometric hip abduction strength test in standing. A fall was defined as an event that results in a person coming to rest unintentionally on the ground or floor or other lower level, not caused by a major intrinsic event or overwhelming hazard (21). A faller was defined as a person who experienced one or more falls during the last 12 months, while a non-faller was defined as a participant who had not fallen during the last 12 months (22). The only inclusion criterion was to be aged 65 years or over. Patients were excluded if they had a major condition that impairs balance such as central nervous system dysfunction (i.e. hemiparesis, myelopathy or cerebellar ataxia), neuromuscular disorders, except distal symmetric peripheral neuropathy (i.e. no myopathy or myasthenia gravis), or vestibular dysfunction. Subjects with severe sepsis, metastatic cancer, angina, or angina-equivalent symptoms with exercise were also excluded. To limit the risks associated with the test, subjects with sores on the plantar surface of the foot, those who had undergone joint replacement within the previous year, had non-consolidated fractures, significant musculoskeletal deformities (i.e. amputation, Charcot-type changes) or lower limb or spinal osteoarthritis were excluded. In order to avoid errors due to a lack of understanding of the test instructions, subjects with moderate or severe dementia (Minimal Mental State Exam (MMSE) < 18) were excluded.

Ethics

The study was approved by the ethical commission in Geneva (CCER - 14-235). All participants signed written informed consent after receiving information about the study and time to make an informed decision regarding participation.

Dynamometer

A calibrated analog dynamometer (SENSIX®, Poitiers, France) that could measure forces between 0 and 667N with a precision of 0.002N was used to measure hip abductor strength (N). It was coupled with the DELSYS® System (Trigno sensor, DELSYS®, INC Boston; MA) that digitalized the analog output (3.3V) with a sampling rate of 1926 Hz and a 16-bit resolution. Use of a HHD to measure lower limb strength has been validated (23,24), and furthermore is not influenced by the experience of the physical therapist (25).

Examiner

The same trained physiotherapist repeated the whole test procedure for all subjects (fallers and non-fallers).

Procedure

The procedure was carried out in 3 parts (*Figure1*): 1) clinical evaluations to characterize participants; 2) the first hip abductor strength test; and 3) the second hip abductor strength test.

The height (m) and weight (kg) of participants was measured first. Then motor function was evaluated using the Timed Up and Go Test (TUG) and the Short Physical Performance Battery (SPPB). Both tests were applied at the time of the inclusion phase, at least one day before the strength test to limit fatigue bias. Following this, the first strength test was carried out. In the pre-tests of this study, we compared the test repetition effect on the performance. We observed non-significant difference between tests highlighting very low learning effect. These previous results justified the absence of learning phase for these fatigable older subjects. To guarantee an identical test condition for all subjects, no pre-test participants were included in the experimental phase. So, none of the participants tried hip abductor strength testing beforehand.

The starting position was standing side-ways-on to the wall (*Figure 2*). The upper arms were by the participant's side and the elbows were flexed to 90° with the hands resting on a treatment table adjusted to the appropriate height. Isometric hip abductor strength was measured using the hand-held dynamometer (HHD). To avoid biases induced by the examiner (sex and strength), the HHD was positioned against the wall, with the examiner hand. The subject's foot didn't touch the ground during the measurements, the knees remained in full extension throughout the test, and the tested hip was in 10° of abduction and in a neutral or slightly extended position (12,16). The center of the HHD was positioned on the lateral malleolus of the tested leg. The neutral hip position and the pressure on lateral malleolus induced a standardized foot position without hip compensation. The subject was instructed to push his/her leg as quickly and as hard as possible towards the dynamometer during five seconds. The tester's job was to hold the dynamometer on the wall, secure the patient, give instructions and give verbal encouragement. Three tests were performed with the right leg, with a minimum 20s of rest between tests with the foot on the ground. The test was repeated on the same day with the same examiner after a break of one to three hours. This time was estimated as being sufficient for a full recovery, but not too long for a change in performance to occur. The examiner

was blind for all measurements because it was the Delsys system that recorded measurements without the need to read the value on the dynamometer screen.

Figure 1: Experimental design.

Data processing

The raw force signals were exported to Matlab® (Mathworks®, Natick, MA, V.8.3.0.532, 2014), which was used for data processing. The signal was low-pass filtered (75ms moving average) to attenuate high-frequency noise. Three dependent variables were extracted: test duration (s), MVIS (N) and RFG (N/s) (13,17). MVIS was defined as the peak value reached within four seconds, and RFG was evaluated during the 50ms after 10% of MVIS was reached (13). Both parameters were normalized to body mass

(26). The mean of the three trials for each test-session was calculated for the MVIS and RFG for each participant.

Figure 2: Set-up for the test.

Statistics

Several analyses were carried out: the number of participants that successfully completed the two test-sessions, the time needed to perform the test and, intra-rater reliability for the whole sample and for each group separately. The statistical analysis included descriptive statistics. Means and standard deviations (SD) are reported for continuous variables. A Mann-Whitney-U test was applied to compare test duration between-group and a Wilcoxon test was applied to compare session 1 and 2. A threshold value of $p < 0.05$ was adopted to rule out the non-significant difference. Intra-rater reliability for hip abductor strength values was determined using the $ICC_{agreement}$ (A,1 type) model as the relative reliability index (27). Interpretation of the $ICC_{agreement}$ was as follows: values >0.75 were considered as “good reliability”, between 0.5 and 0.75 as “moderate reliability” and < 0.5 as “poor reliability” (28). In addition, we calculated confidence intervals (CI 95%) through bootstrapping (5000 resamples, bias corrected and accelerated percentile method) (13). If the $ICC_{agreement}$ value was above 0.50, the measurement error value using the standard error of measurement was computed: $(SEM = ST\sqrt{1 - ICC})$ with ST being the standard

deviation of all the trials of both sessions. The smallest detectable difference (SDD) was also calculated: ($SDD = 1.96 * SEM * \sqrt{2}$). The SDD was normalized by the mean and expressed as percentage. Absolute reliability was investigated using Bland and Altman analysis to determine between-session agreement of the strength measurements (16,27). The 95% limits of agreement (LOA 95%) represent 2 standard deviations (SD) above and below the mean difference (bias) between sessions.

RESULTS

Clinical characteristics

The number of potentially eligible subjects was 48. Seven subjects refused to participate (without explanation: 2, fatigue: 1, too many other exams at hospital: 2 and 2 subjects refused to sign informed consent), and 9 subjects were excluded by the investigator (too early discharge: 2, MMSE too low: 4, diagnosis: 2, poor language comprehension: 1). Finally, thirty-two older persons were included (16 fallers and 16 non-fallers). The characteristics of the whole study population as well as the characteristics per sub-group are presented in Table 1.

Table 1. Description in mean (±SD) of all patients and the 2 subgroups (fallers and non-fallers) recruited for the hip abduction test in standing position.

Variables	All (N=32)	Fallers (N=16)	Non-fallers (N=16)
Age (years)	83.03 ± 7.78	86.94 ± 6.23	79.12 ± 7.33
Sex (%)	Men: 34% / Women: 66%	Men: 44% / Women: 56%	Men: 25% / women: 75%
Weight (kg)	63.88 ± 12.47	59.93 ± 11.87	67.83 ± 12.13
Height (m)	1.64 ± 0.08	1.62 ± 0.08	1.66 ± 0.07
BMI (kg/m²)	23.62 ± 3.78	22.61 ± 3.65	24.62 ± 3.74
Leg dominance (%)	Left: 6% / Right: 94%	Left: 6% / Right: 94%	Left: 6% / Right: 94%
Number of falls (last 12 months)	N/A	3.25±1.61	N/A
Walking aid (%)	Yes: 31% / No:69%	Yes: 56% / No: 44%	Yes: 6% / No: 94%
TUG (s)	15.90 ± 9.99	22.46 ± 10.33	9.34 ± 2.70
SPPB (score /12)	8.59 ± 3.50	5.88 ± 2.68	11.31 ± 1.54
MMSE (score /30)	27.06 ± 3.17	25.31 ± 3.40	28.81 ± 1.64

BMI: body mass index, TUG: timed up and go, SPPB: short physical performance battery, MMSE: mini mental state examination

Feasibility of the tests

All 32 participants successfully completed both hip abductor strength-tests. Assessment-time was less than four minutes (mean 2.47 ± 0.49 minutes per leg), not-including set-up time (<1 min). For whole group, we observed non-significant difference between sessions (session 1: 166.87 ± 43.06s vs. session 2: 167.81 ± 55.40s, non-significant p value - NS). The comparison of duration was not significant between sub-groups for either session (session 1: non-fallers: 165.00 ± 45.17s vs. fallers: 168.75 ± 42.25s; session 2: non-fallers: 151.88 ± 31.88s vs. fallers: 183.75 ± 42.25s, NS). For each subgroup, the comparison between sessions was NS.

Intra-rater reliability analysis

Table 2 shows the intra-rater reliability of the test using the ICC_{agreement}, SEM and SDD. The total sample MVIS and RFG measures were reliable, with ICC_{agreements} above 0.75 although the ICC_{greement} = 0.93 [0.87 – 0.97] for the RFG (fallers: 0.93 [0.84 – 0.98] and non-fallers: 0.88 [0.75 – 0.96]) was lower than that of the MVIS with ICC_{agreement} = 0.98 [0.95 – 0.99] (fallers: 0.94 [0.89 – 0.98] and non-fallers: 0.98 [0.95 – 1.00]). This was also found for the sub-groups. The SEM values for MVIS ranged from 0.11 N.kg⁻¹ (fallers) to 0.08 N.kg⁻¹ (non-fallers). For RFG, SEM values ranged from 1.12 N.kg⁻¹.s⁻¹ (fallers) to 1.34 N.kg⁻¹.s⁻¹ (non-fallers). The SDD was 32.7% for MVIS and 48.8% for RFG in fallers. SDD values were lower in non-fallers (MVIS: 20.3%, RFG: 41.2%).

Table 2: ICC_{agreement} (A-1), SEM and SDD (%) values for MVIS and RFG parameters of all patients and the 2 subgroups (fallers and non-fallers) recruited for the hip abduction test in standing position.

Variables		ICC _{agreement} [95%CI]	ICC _{agreement} interpretation	SEM (unit of measure)	SDD (%)
All (N=32)	MVIS	0.98 [0.95 – 0.99]	Good	0.09	25.2%
	RFG	0.93 [0.87 – 0.97]	Good	1.24	44.3%
Fallers (N=16)	MVIS	0.94 [0.89 – 0.98]	Good	0.11	32.7%
	RFG	0.93 [0.84 – 0.98]	Good	1.12	48.8%
Non-fallers (N=16)	MVIS	0.98 [0.95 – 1.00]	Good	0.08	20.3%
	RFG	0.88 [0.75 – 0.96]	Good	1.34	41.2%

ICC_{agreement}: intra-class correlation coefficient, SEM: standard of error measurement, SDD: smallest detectable difference, MVIS: maximal voluntary isometric strength, RFG: rate of force generation

The MVIS and RFG values did not differ across test-sessions, showing that no learning effect occurred (no systematic effect) (Figure 3). The mean differences were larger and the LOA smaller in the fallers than in the non-fallers (Table 3).

Comparison of hip MVIS and RFG between test-sessions in fallers and non-fallers

There were non-significant differences in hip abductor strength between sessions for both parameters (MVIS and RFG) for the whole group and both sub-groups. MVIS was significantly lower in the fallers compared to the non-fallers in both the first (fallers: $0.94 \pm 0.32 \text{ N.kg}^{-1}$ vs. non-fallers: $1.22 \pm 0.42 \text{ N.kg}^{-1}$, $p=0.0012$) and second tests (fallers: $0.93 \pm 0.28 \text{ N.kg}^{-1}$ vs. non-fallers: $1.21 \pm 0.46 \text{ N.kg}^{-1}$, $p=0.0126$). A similar result was found for the RFG in both the first (fallers: $6.75 \pm 3.12 \text{ N.kg}^{-1}.\text{s}^{-1}$ vs. non-fallers $9.10 \pm 2.75 \text{ N.kg}^{-1}.\text{s}^{-1}$, $p=0.0022$) and second tests (fallers: $6.08 \pm 3.24 \text{ N.kg}^{-1}.\text{s}^{-1}$ vs. non-fallers $9.31 \pm 3.23 \text{ N.kg}^{-1}.\text{s}^{-1}$, $p=0.0001$). In the scatterplots (Figure 3, plots A and B), the values of the open dots (fallers) are lower (less strength) than the closed dots (non-fallers).

Table 3: Bland and Altman limits of agreement between sessions for MVIS and RFG parameters of all patients and the 2 subgroups (fallers and non-fallers) recruited for the hip abduction test in standing position.

	Variables	Mean ($\pm$ SD) difference between sessions (bias)	Lower LOA to Upper LOA
All (N=32)	MVIS (N.kg^{-1})	0.007 ± 0.13	-0.38 to 0.40
	RFG ($\text{N.kg}^{-1}.\text{s}^{-1}$)	0.23 ± 1.75	-3.01 to 3.48
Fallers (N=16)	MVIS (N.kg^{-1})	0.008 ± 0.16	-0.28 to 0.29
	RFG ($\text{N.kg}^{-1}.\text{s}^{-1}$)	0.67 ± 1.48	-2.42 to 3.76
Non-fallers (N=16)	MVIS (N.kg^{-1})	0.006 ± 0.12	-0.48 to 0.44
	RFG ($\text{N.kg}^{-1}.\text{s}^{-1}$)	-0.21 ± 1.93	-3.04 to 2.63

SD: standard deviation, LOA: limits of agreement, MVIS: maximal voluntary isometric strength, RFG: rate of force generation

Figure 3: Correlations between sessions 1 and 2 for all participants and both parameters (MVIS and RFG) are represented in plots A and B respectively. The Bland and Altman plot analysis for MVIS is represented in plot C and the RFG in plot D.

DISCUSSION

The aim of this work was to evaluate the reliability of a functional and user-friendly test to measure hip abductor strength in older people. This is particularly important since hip abductor strength is known to be associated with a risk of falls (4). The results showed that testing hip strength in a standing position is feasible and rapid. All participants successfully completed the test in less than four minutes. Moreover, intra-rater reliability was good for the total sample as well as for both the faller and non-faller subgroups. As testing hip strength in a standing position is feasible, reliable, and better reflects real life situations than strength tests in supine or side-lying positions, we recommend this position for clinical practice.

The test-retest reliability for MVIS found in the present study was similar to that found by Wang et al. in supine position (29). They evaluated the reliability of strength measurements using a HHD in 8 lower limb muscle groups in community-dwelling, older fallers. However, this study included only fallers without a comparison with non-fallers. In our study, ICC values for MVIS were better in non-fallers than fallers, while the intra-rater reliability for RFG was lower for the non-fallers than for fallers. Nevertheless, ICC values remained above 0.75 which is a good reliability and an appropriate threshold value for clinical use (28). Our results and those of Wang et al. (29) showed that hip abductor strength can be measured reliably in older subjects at risk of falls both in supine and standing position. Widler et al. assessed intra-rater reliability in young adults (12). They compared intra-rater reliability in standing (ICC=0.88), side-lying (ICC=0.90), and supine positions (ICC=0.82). They found the highest ICC in side-lying, although reliability was also good in the two other positions. The level of test-retest reliability found in the present study was higher than in the study by Widler et al. (12). This difference might be due to more heterogenic strength values for our study sample, but also due to differences in the test position (21). In the present study both forearms of the participants were supported on a high treatment table, while in the study by Widler et al. subjects stabilized themselves with only one hand (12). The importance of stabilization on reliability has been confirmed by a recent study (16). Awwad et al. (16) who assessed hip abductor strength in older adults in a standing position showed values close to our results. In contrast to our study their participants stabilized the pelvis against a wall and participants indicated to perceive fatigue in the opposite hip to that being tested (16). This was probably due to the fact that standing leg was necessary to stabilize the pelvis in this position. Awwad et al. (16) further showed an increased intra-subjects variability in strength values recorded in the standing position compared to the supine position (in which the pelvis was also stabilized against the wall). In our study the pelvis was not stabilized against a wall. Nevertheless, we found similar ICC's to Awwad et al.

and none of the patients complained of fatigue in the supporting leg. These facts support the use of the testing position described in the current study. In addition, our test has the advantage of being closer to different motor functions during which the pelvis is never stabilized.

Another explanation for the discrepancy between the study of Widler et al. (12) and our results could be the differences in the subjects' ages. The greater strength of young subjects induced certainly greater difficulties to avoid compensations (15). As the muscle strength in fallers was lower than in non-fallers, it is likely that fallers compensated less which might have increased reliability. In the supine position with belt-resisted dynamometer fixation during strength testing, Kramer et al. found higher test-retest reliability in older subjects (ICC>0.98) than in young adults (ICC>0.92) (15). The authors attributed this ICC difference to the compensation during test but also to higher between-subject strength variation in older adults. In standing position, the intra-rater reliability of MVIS measured in older adults in our study was higher (ICC>0.94) than the intra-rater reliability of MVIS in young healthy subjects of the study of Widler et al. (ICC>0.88) (12). The results of these studies corroborate the results of Kramer et al. who tested older and young adults in a side-lying position (15). Among the study population fallers were eight years older than non-fallers. However, the ICC of hip abductor MVIS were similar in fallers and in non-fallers whereas higher ICC's were found for the RFG measures in fallers (ICC=0.93) than in non-fallers (ICC=0.88). It might be that the RFG is more influenced by age than the MVIS and thus induces more heterogeneity and consequently a higher ICC values in RFG parameter. It is well known that movement slows down with advancing age which is thought to be partially strategic in that older adults give emphasis to movement accuracy at the cost of movement speed (30). In addition, slower information processing due to an increase in neural noise and other synaptic changes may also affect the RFG (30). Nevertheless, we cannot exclude that several other unknown factors playing a role in these differences.

In both the present study and the study by Widler et al. (12), the HHD was fixed. Several studies have shown that when a HHD is fixed externally, subjects generate higher magnitudes of peak force, and test-retest reliability is higher than for manual, examiner resisted HHD measurements (15,31). In our study, the HHD was fixed to the wall, thus the procedure was standardized, and no bias was created by variations in rater-force. Moreover, the procedure is faster when the HDD is fixed externally.

Comparison of the strength parameters showed that reliability was higher for MVIS than RFG. To our knowledge, no other study has evaluated the test-retest reliability of the RFG of the hip abductors in a standing position, despite the fact that RFG is an important parameter relating to dynamic balance control in older subjects (4,17,18). Although the reliability of RFG was lower than MVIS, it was still

good ($ICC > 0.88$). Thus, according to clinical recommendations by Portney and Watkins (28), this parameter can be used to evaluate hip abductor strength in clinical practice. We believe it is relevant to assess both MVIS and RFG since they are not correlated (20), and both are associated with an increased risk of falls in older women (4). In clinical practice, the dynamometer usually directly shows MVIS, but not RFG. Thus, in the future, it would be useful to develop HHD tools with RFG values directly displayed on the screen facilitating the use by clinicians.

ICCs are highly dependent on the heterogeneity of the study sample while SEM, SDD and LOA are more appropriate to evaluate changes over time (32). SEM values are also interesting to clinicians in terms of decision making, since they describe errors in the same units of measurement, and can be used to calculate SDD between two measurements (33). In standing, the SEM for MVIS was very similar between fallers and non-fallers, however, the SDD was higher in fallers (respectively 32.7% and 20.3%). In whole group, the SDD of RFG was higher (44.3%) than the SDD of MVIS (25.2%). Mentiplay et al. showed the same results for young subjects measured in a supine position with an SDD of RFG of 34.65% and an SDD of MVIS of 20.23% (23). The SDDs found in the present study were lower than the ones in our previous study that measured hip abductor strength in older adults in a side-lying position (RFG: 51%, MVIS: 32%) (13). Thus, we recommend in clinical practice testing hip abductor strength in older persons at risk of falls in a standing position. Similarly to Widler et al. (12) and Awwad et al. (16), who evaluated maximal hip strength value in a standing position, Bland and Altman analysis for intra-rater reliability showed no systematic session effect for the MVIS. This suggests that *no learning effect* or fatigue occurred in the patients or raters and confirms the quality of our procedure without previous practical trials for this parameter. Nevertheless, Bland and Altman analysis showed somewhat higher mean between sessions differences for RFG, especially for the fallers, indicating a small bias, i.e. a small learning effect between sessions. The greater mean differences between session for RFG than for MVIS have also been mentioned by Mentiplay et al. (23). The good intra-rater reliability and the small systematic effect which has previously been observed for MVIS in older people (16), the lower SDD observed in older persons at risk of falls in a standing position compared to the ones of older persons in a side-lying (13) and the feasibility highlight the interest of the clinical abductor test in standing position.

It should be noted that the present study only evaluated intra-rater reliability. In addition, the test position of participants was only verified visually without the use of precise sensors. However, this reflects clinical practice. In addition, the rater was not blinded to the faller or non-faller status of the participant. In this study, before tests, the subjects haven't had a separate familiarization session. We decided for this procedure according to our pilot test as well as to avoid additional fatigue for these

older subjects. Indeed, we think that fatigue is a more important bias than the lack of a familiarization session. However, we believe that the reliability could be improved for patients with low fatigue and who have had a practice trial, especially for RFG parameter. The MVIS and RFG parameters were assessed in the same trial, in order to avoid too many repetitions for the participants. This implicated a more complex test instruction (push hard and fast) which may have been difficult for some participants to achieve. Testing MVIS and RFG in two separate trials may produce more accurate results for the assessment of strength in an older population.

This study focused on immediate isometric hip abductor strength without considering endurance, which is an important parameter for activities of daily living. Moreover, it is well known that type 1 muscle fibers tend to degenerate in older people (34,35). Van Cant et al. (35) found good test-retest reliability for both isometric and isotonic hip abductor endurance in a side-lying position in young adults. Therefore, future studies should determine the reliability of testing both maximal isometric strength and endurance of the hip abductors in a standing position in older subjects.

CONCLUSION

Assessment of hip abductor strength (MVIS and RFG) using a HHD in a standing position is feasible and reliable in older people at risk of falls. The test is easy to carry out in clinical practice.

The intra-rater reliability of abductor strength test with HDD seems associated to study population (age, pathology, ...), testing position and testing procedure. Our results showed that hip abductor strength in older persons can reliably be measured in a standing position with the dynamometer fixed against a wall, pelvis frontal movement not restricted and an upper limb support on a treatment table to secure the patient. Moreover, assessment in a standing position is more functional than in supine or side-lying and provides a better evaluation of real-life activities in which falls occur. We thus recommend assessment of hip abductor strength in a standing position.

LIST OF ABBREVIATIONS

ICC_{agreement}: Intraclass Correlation Coefficient Agreement
 LOA: Limits of agreement
 HHD: hand-held dynamometer
 MVIS: Maximum Voluntary Isometric Strength
 RFG: Rate of Force Generation
 SEM: Standard Error of Measurement
 SDD: Smallest detectable difference
 TUG: Time Up and Go Test
 SPPB: Short Physical Performance Battery
 MMSE: Minimal Mental State Exam

REFERENCES

1. ActiveAgeing_ok - WHO_NMH_NPH_02.8.pdf [Internet]. [cited 2017 Apr 5]. Available from: http://apps.who.int/iris/bitstream/10665/67215/1/WHO_NMH_NPH_02.8.pdf
2. Boyé NDA, Mattace-Raso FUS, Van der Velde N, Van Lieshout EMM, De Vries OJ, Hartholt KA, et al. Circumstances leading to injurious falls in older men and women in the Netherlands. *Injury*. 2014;45:1224–30.
3. Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson LM, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2012;CD007146.
4. Morcelli MH, Rossi DM, Karuka AH, Crozara LF, Hallal CZ, Marques NR, et al. Peak torque, reaction time, and rate of torque development of hip abductors and adductors of older women. *Physiother Theory Pract*. 2016;32:45–52.
5. Arvin M, van Dieën JH, Faber GS, Pijnappels M, Hoozemans MJM, Verschueren SMP. Hip abductor neuromuscular capacity: A limiting factor in mediolateral balance control in older adults? *Clin Biomech Bristol Avon*. 2016;37:27–33.
6. Allet L, Kim H, Ashton-Miller J, De Mott T, Richardson J. Step length after discrete perturbation predicts accidental falls and fall-related injury in elderly people with a range of peripheral neuropathy. *J Diabetes Complications*. 2014;28:79–84.
7. Arvin M, Hoozemans MJM, Burger BJ, Rispen SM, Verschueren SMP, van Dieën JH, et al. Effects of hip abductor muscle fatigue on gait control and hip position sense in healthy older adults. *Gait Posture*. 2015;42:545–9.
8. Laroche DP, Cook SB, Mackala K. Strength asymmetry increases gait asymmetry and variability in older women. *Med Sci Sports Exerc*. 2012;44:2172–81.
9. Terrier P, Reynard F. Effect of age on the variability and stability of gait: a cross-sectional treadmill study in healthy individuals between 20 and 69 years of age. *Gait Posture*. 2015;41:170–4.
10. Reynard F, Vuadens P, Deriaz O, Terrier P. Could Local Dynamic Stability Serve as an Early Predictor of Falls in Patients with Moderate Neurological Gait Disorders? A Reliability and Comparison Study in Healthy Individuals and in Patients with Paresis of the Lower Extremities. *PLoS ONE* [Internet]. 2014 [cited 2017 Apr 5];9. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4065053/>
11. Addison O, Inacio M, Bair W-N, Beamer BA, Ryan AS, Rogers MW. Role of Hip Abductor Muscle Composition and Torque in Protective Stepping for Lateral Balance Recovery in Older Adults. *Arch Phys Med Rehabil*. 2016;
12. Widler KS, Glatthorn JF, Bizzini M, Impellizzeri FM, Munzinger U, Leunig M, et al. Assessment of hip abductor muscle strength. A validity and reliability study. *J Bone Joint Surg Am*. 2009;91:2666–72.
13. Gafner S, Bastiaenen CHG, Terrier P, Punt I, Ferrari S, Gold G, et al. Evaluation of hip abductor and adductor strength in the elderly: a reliability study. *Eur Rev Aging Phys Act Off J Eur Group Res Elder Phys Act*. 2017;14:5.

14. Nadler SF, DePrince ML, Hauesien N, Malanga GA, Stitik TP, Price E. Portable dynamometer anchoring station for measuring strength of the hip extensors and abductors. *Arch Phys Med Rehabil*. 2000;81:1072–6.
15. Kramer JF, Vaz MD, Vandervoort AA. Reliability of isometric hip abductor torques during examiner- and belt-resisted tests. *J Gerontol*. 1991;46:M47–51.
16. Awwad DH, Buckley JD, Thomson RL, O'Connor M, Carbone TA, Chehade MJ. Testing the Hip Abductor Muscle Strength of Older Persons Using a Handheld Dynamometer. *Geriatr Orthop Surg Rehabil*. 2017;8:166–72.
17. Allet L, Kim H, Ashton-Miller J, De Mott T, Richardson JK. Frontal plane hip and ankle sensorimotor function, not age, predicts unipedal stance time. *Muscle Nerve*. 2012;45:578–85.
18. Allet L, Kim H, Ashton-Miller JA, Richardson JK. Which lower limb frontal plane sensory and motor functions predict gait speed and efficiency on uneven surfaces in older persons with diabetic neuropathy? *PM R*. 2012;4:726–33.
19. Gruber M, Gollhofer A. Impact of sensorimotor training on the rate of force development and neural activation. *Eur J Appl Physiol*. 2004;92:98–105.
20. Chourasia AO, Buhr KA, Rabago DP, Kijowski R, Irwin CB, Sesto ME. Effect of Lateral Epicondylitis on Grip Force Development. *J Hand Ther*. 2012;25:27–37.
21. Tinetti ME, Speechley M, Ginter SF. Risk Factors for Falls among Elderly Persons Living in the Community. *N Engl J Med*. 1988;319:1701–7.
22. Howcroft J, Lemaire ED, Kofman J, McIlroy WE. Elderly fall risk prediction using static posturography. *PLoS One*. 2017;12:e0172398.
23. Mentiplay BF, Perraton LG, Bower KJ, Adair B, Pua Y-H, Williams GP, et al. Assessment of Lower Limb Muscle Strength and Power Using Hand-Held and Fixed Dynamometry: A Reliability and Validity Study. *PLoS One*. 2015;10:e0140822.
24. Charlton PC, Mentiplay BF, Grimaldi A, Pua Y-H, Clark RA. The reliability of a maximal isometric hip strength and simultaneous surface EMG screening protocol in elite, junior rugby league athletes. *J Sci Med Sport*. 2017;20:139–45.
25. Larson CA, Tezak WD, Malley MS, Thornton W. Assessment of postural muscle strength in sitting: reliability of measures obtained with hand-held dynamometry in individuals with spinal cord injury. *J Neurol Phys Ther JNPT*. 2010;34:24–31.
26. Hsu A-L, Tang P-F, Jan M-H. Test-retest reliability of isokinetic muscle strength of the lower extremities in patients with stroke. *Arch Phys Med Rehabil*. 2002;83:1130–7.
27. Seko T, Kumamoto T, Miura S, Kobayashi T, Takahashi Y, Kaneko R, et al. Measuring seated hip extensor strength using a handheld dynamometer: an examination of the reliability and validity of the protocol. *J Phys Ther Sci*. 2015;27:2179–82.
28. Portney L, Watkins M. *Foundations of clinical research: applications to practice*. 3rd ed. Pearson/Prentice Hall: Upper Saddle River, NJ; 2009.
29. Wang C-Y, Olson SL, Protas EJ. Test-retest strength reliability: hand-held dynamometry in community-dwelling elderly fallers. *Arch Phys Med Rehabil*. 2002;83:811–5.
30. Seidler RD, Bernard JA, Burutolu TB, Fling BW, Gordon MT, Gwin JT, et al. Motor control and aging: links to age-related brain structural, functional, and biochemical effects. *Neurosci Biobehav Rev*. 2010;34:721–33.
31. Schache MB, McClelland JA, Webster KE. Reliability of measuring hip abductor strength following total knee arthroplasty using a hand-held dynamometer. *Disabil Rehabil*. 2016;38:597–600.
32. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033–9.
33. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. *Phys Ther*. 1993;73:386–95.
34. Reid KF, Pasha E, Doros G, Clark DJ, Patten C, Phillips EM, et al. Longitudinal decline of lower extremity muscle power in healthy and mobility-limited older adults: influence of muscle mass, strength, composition, neuromuscular activation and single fiber contractile properties. *Eur J Appl Physiol*. 2014;114:29–39.
35. Van Cant J, Dumont G, Pitance L, Demoulin C, Feipel V. Test-retest reliability of two clinical tests for the assessment of hip abductor endurance in healthy females. *Int J Sports Phys Ther*. 2016;11:24–33.

CHAPTER 4

Hip muscle and hand grip strength
to differentiate between older fallers and non-fallers;
a cross-sectional validity study

Simone C. Gafner, Caroline H. G. Bastiaenen, Serge Ferrari, Gabriel Gold,
Philippe Terrier, Roger Hilfiker, Lara Allet

Clinical interventions in aging. 2018;13:1-8.
DOI: 10.2147/CIA.S146834

ABSTRACT

Background:

Hip muscle weakness in older persons seems to be an influencing factor of falls. Currently it is unclear which muscles out of the hip muscle group play an important role in older persons. A validating process in the measurement regarding muscles strength related to falls is necessary before answering that question.

Objective:

We aimed to investigate which hip muscles strengths group shows an acceptable level of distinction between older adult fallers and non-fallers compared to a pre-defined external criterion regarding falling. Secondly, we aimed to compare the same outcomes and questions for hand-grip strength in relation to the same external criterion.

Design:

Cross-sectional validity study

Methods:

The maximum voluntary isometric strength (MVIS) and the rate of force generation (RFG) of hip abductors (ABD), adductors (ADD), internal and external rotators (IR and ER), extensors (EXT) and flexors (FLEX) were measured with a dynamometer fixed to a custom-made frame as well as hand-grip strength with a Martin Vigorimeter in 60 older persons aged over 65 years (38 females, 22 males).

Results:

The area under the curve (AUC) and the results of a random forest Mean Decreased Gini Index show, that of all assessed parameters hip ABD MVIS is the parameter with the highest discriminative value regarding the chosen external criterion in older persons (AUC ABD MVIS 0.825, confidence interval 0.712 to 0.938).

Conclusion:

Results indicate that ABD MVIS is a useful measure to distinguish between older adult fallers and non-fallers regarding the chosen external criterion.

Keywords: hip muscle strength, accidental falls, older adults, hand grip strength, measurement study

INTRODUCTION

High fall prevalence, especially multiple falls, increases the risk of independency loss (1), admission to nursing homes (2), reduction in quality of life (3) and mortality (4) in older persons. The World Health Organization (WHO) reports a worldwide fall rate of approximately 28-35% for persons aged over 65 years. They also report a yearly fall rate of approximately 30-50% of people living in long-term care institutions whereas 40% of these persons show recurrent falls (5). For Switzerland, the federal statistical office (FSO) reports similar yearly fall rates of 25% for persons over 65 years living at home (6) and 39% for persons in long-term care institutions (7).

Lower limb weakness in older persons is assumed to be an important influencing factor of falls (8) and it's preservation potentially contributes to the maintenance of independency (9).

There are several reasons for the increasing interest for the relationship between hip muscle strength and fall risk of older persons. Approximately 7% of older persons suffer from peripheral neuropathy (10), which is a predominantly distal dysfunction linked to an increased fall risk (11). The preservation of proximal lower limb strength might enable older persons to compensate for distal nerve function losses (12). Next, older persons tend to show a more accentuated hip strategy whereas younger counterparts use rather an ankle strategy to keep their balance (13). Finally, older persons, in particular fallers, have a reduced medio-lateral stability while standing (14) and increased medio-lateral body motion during dynamic balance recovery in forward stepping tasks (15). This medio-lateral control is mainly assured through proximal (hip) frontal plane muscle strength. Therefore hip muscles strength might be an important parameter to include in fall risk assessment and can have a role in the prevention and rehabilitation in an older population at risk of falls. Hip muscles strength is an easily assessable parameter and most importantly easy to target in physical therapy. To contribute to goal-setting for treatment, insight in diagnostic accuracy of measurement of strength of hip muscles to apply in an older population at risk of falling seems to be important.

To our best knowledge no study exists that simultaneously assessed and compared all six hip muscle groups to investigate its diagnostic values to discriminate between older adults with and without fall history (as an assumed external criterion). Fall history is known as a risk factor for future falls (16). However it's assessment doesn't give us a directly targetable parameter for the physical therapy treatment.

In addition, a separate comparison between hip muscle strength and hand-grip strength with the same external criterion is not yet investigated. At the moment, hand-grip strength is assumed to be a parameter related to whole body strength and fall risk (17).

For this first step of a validation of a new measurement tool, the comparison of currently used assessments with the newly, to be developed fall risk assessment tool is essential to get an idea about it's discriminating ability between older adults fallers and non-fallers. Future studies will have to prospectively assess the hip muscle strength group with the highest diagnostic accuracy.

Our aims are:

- a) To identify the level of distinction of the different hip muscle group strengths, evaluated by the area under the curve (AUC) >0.7, compared to a pre-defined external criterion regarding falling by measuring maximal voluntary isometric strength (MVIS) and rate of force generation (RFG).
- b) To compare the outcomes of hip muscle strength with the hand-grip strength compared to the same external criterion.

We hypothesized that a) within the group of hip muscles, the hip frontal plane muscle groups, essentially hip abductors strengths, show the highest level of AUC > 0.7 and b) that the outcomes of frontal plane hip muscle strength expressed in AUC will have higher values than hand-grip strength compared to the same external criterion.

METHODS

We examined the concurrent validity of the different hip muscles strengths groups as well as hand-grip strength compared to older participants' history of falling as external criterion. The data collection was performed cross-sectionally. The study is designed according to the guidelines of the COSMIN TAXONOMY regarding criterion validity (18).

Participants

In a limited period of time, from March to October 2015, sixty participants were simultaneously recruited in the Geriatric Department of the University Hospital of Geneva as well as in an ambulatory setting. The two different settings guaranteed our heterogeneous sample needed for a validity study. Participants admitted to physical therapy for musculoskeletal or vascular disorders were consecutively checked for the in- and exclusion criteria (see below) and thereafter asked if they agree to participate in this study (Figure 1, flow chart of participants recruitment). The protocol was approved by the ethics committee of Geneva (commission cantonale d'éthique de la recherche (CCER) 14-235). All participants received written and oral information and signed informed consent.

Participants were included when over 65 years old, with a body mass index (BMI) between 17 and 35 and able to walk household distances. They were excluded in case of a positive history or evidence of any significant central nervous system dysfunction, neuromuscular disorders except a distal symmetric peripheral neuropathy, evidence of vestibular dysfunction, moderate or severe dementia (Minimal Mental State Exam (MMSE) < 18), a fracture of the lower limb and/ or joint replacement within the previous year and any lower limb or back pain that adversely affects the strength tests.

Measures

External Criterion

The external criterion was defined as a positive fall history within the previous 12 months. Participants reporting one or more falls in the last year were classified as fallers (19). Persons that didn't fall were classified as non-fallers (19). An event which results in a person coming to rest inadvertently on the ground or floor or other lower level was defined as a fall (5). The fall history was chosen as external criterion in this study because of the ultimate importance of this discriminating factor within our theoretical construct of contributing factors regarding fall risk in older persons. Identifying a suitable criterion that is considered to represent the true state of the construct of interest is a challenge. As the construct of interest is discriminating older people in groups based on a directly targetable parameter related to a high risk of falling or not having a high risk of falling; fall history is a suitable external criterion.

Index measures

Hip strength was measured in Newton (N) using an analog dynamometer (SENSIX®, Poitiers, France) with a measuring range from 0 to 667 N and a precision of 0.002N. The DELSYS System (Trigno sensor, DELSYS, INC Boston; MA) coupled with the calibrated dynamometer digitalized the output (3.3V) with a sampling rate of 1926 Hz and a resolution of 16 bit (20). To avoid bias through testing persons the dynamometer was fixed to a frame for all strength tests. Hip MVIS (N/kg) and RFG (N/s/kg) was assessed in all anatomic planes. Subjects pushed as fast and as hard as possible against the dynamometer (21), held the pressure for three seconds and relaxed (12). Participants repeated all tests three times per leg with a break of one minute between each trial. The mean of the six trials was retained for further analyses. The starting position for each test was supported by the examiner in one centimeter (cm) distance of the dynamometer. The tester counted loudly from one to three to give the starting signal. Strong verbal encouragement was given during all

measurements. The testing positions for the hip strength tests as well as the hand-grip strength are described in table 1.

Characteristics of the population

The collected variables were age, gender, weight, BMI, the highest educational level achieved (27), reported dominant hand, the level of concern about falling assessed with the Falls Efficacy Scale International (FES-I) and cognitive impairments with the Mini Mental State Examination Questionnaire (MMSE).

Figure 1: Flow chart of the recruitment of the participants. MMSE; Mini mental state examination

Table 1: Hip strength test position and dynamometer placement as well as hand-grip strength measurement position with the Martin Vigorimeter

Strength tests	Test position	References for test positions	Dynamometer position	Special aspects to consider
Hip flexors (FLEX)	Supine with knee and hip in 90° flexion. Neutral position according to rotations.	Mendis et al (22)	5cm cranial to the knee cap	
Hip extensors (EXT)	Standing with the hip in neutral position according to rotations.	Kollock et al (23)	In the middle of the line between the os calcaneum and the knee joint gap on the posterior side of the calf.	The knee had to stay extended and the trunk in a neutral, upright position. If participants couldn't perform the test in the correct position, the EXT was reported as missing value
Hip internal Rotators (IR)	Sitting position with the knee and hip flexed at 90°	Thorborg et al (24)	5 cm cranial of the external malleolus	Special attention was paid that participants didn't execute an abduction movement of the hip
Hip external Rotators (ER)	Sitting position with the knee and hip flexed at 90°	Thorborg et al (24)	5 cm cranial to the internal malleolus	Special attention was paid that participants didn't execute an adduction movement of the hip
Hip abductors (ABD)	Side-lying position with the upper leg in a 10° abducted position. The lower leg was in 45° of knee and hip flexion for stabilization	Krause et al (25)	5 cm cranial to the malleolus externus	The tested leg was extended, with the hip in a neutral or slightly extended position
Hip adductors (ADD)	Side-lying position with the tested lower leg in a neutral position. The participants upper leg was placed on a pillow in a 90° hip and knee flexion	Thorborg et al (24)	5 cm cranial to the medial malleolus	The tested leg was extended, with the hip in a neutral or slightly extended position
Hand-grip strength	Sitting on a chair with the elbow in a 90° flexion position. The forearm was in a neutral position and the wrist in 0 to 30° extension.	Desrosiers et al (26)	Participants were asked to squeeze the rubber bulb of the Martin Vigorimeter as hard as possible and to relax after three seconds.	

Falls Efficacy Scale International (FES-I)

Fear of falling is known to be an influencing factor of falls (28). The Falls Efficacy Scale International (FES-I) is a validated questionnaire including 16 questions assessing the level of concern about falling of older persons (29).

Mini Mental Status Examination (MMSE)

Cognitive impairment influences the fall risk in older adults (30). The MMSE is a widely used (30) and validated assessment tool for cognitive impairments in older persons consisting of 11 questions (31).

Data collection

After giving informed consent we collected participants socio-demographic data; age, gender, weight, BMI by extraction of the clinical database and the highest educational level achieved, the dominant hand and the external criterion fall history by interview. To further investigate the characteristics of our study population a trained physical therapist assessed the MMSE and the FES-I. Then subjects' hip and hand-grip strength was tested in the same way.

Sample size

We expected to find AUC values between good (0.8) to excellent (0.9) (32). To test the null hypothesis H_0 : AUC = 0.7 (H_1 : AUC $\neq$ 0.7) with an expected AUC of 0.85, a sample of n=48 is appropriate to achieve a significance level of 5% and a power of 80% (33). In order to compensate for an expected drop-out rate of 20 % we aimed to totally include 60 participants.

Statistics

Statistics were performed with Stata version 14.1 (College Station, Texas 77845, USA) and the statistical environment R (34). Descriptive data (gender, age, weight, BMI, MMSE, FES-I) as well as the outcome data for strength are described in Table 2 and Table 3. Missing data were not imputed in our statistics.

Data Processing

The raw force signals were exported to Matlab (Mathworks, Nathick, MA, version 8.3.0.532). A low-pass filter (75 milliseconds (ms) moving average) was applied to attenuate high-frequency noise. The MVIS was defined as the peak value reached within zero to four seconds. The RFG was evaluated over 50ms after 10% of the MVIS was reached. All hip strength values are normalized for participants' weight.

Main analyses

The diagnostic value of the different hip muscle groups and the hand-grip strength based on the gold standard (history taking of falling) was investigated with the area under the curve (AUC) of the receiver operating characteristics (Table 3) based on a univariable logistic regression. An AUC > 0.7 is a moderately accurate level for a diagnostic test (35) which we judged as an acceptable level for the clinical use of this diagnostic measurement tool. Participants with missing strength data were not included in the calculations of the AUC (see results).

The variable importance of the strength tests was investigated with a random forest method with a consensus prediction reached for each observation by averaging the results of many individual recursive partitioning tree models (36). A training set was drawn from the original data using bootstrap with replacement and a classification tree was built. This was repeated 50'000 times and the final classification is the one that appears the most. The Mean Decrease Gini Index (DGI), which indicates the quality of a split for every variable by the means of the Gini Index, was calculated with the package randomForest (37) within the statistical environment R (34).

RESULTS

Sixty participants with a mean age of 82.0 years and a standard deviation (SD) of 6.61 were finally included in this study (Figure 1, flow chart of the patient recruitment). Due to missing data (EXT n= 3, IR n= 2, ER n= 2) of a total of four participants the analyses of the AUC and DGI were performed including 56 participants. Most of our participants were right handed (N=53) and had a medium educational level (terminated apprenticeship) (N= 34). Twenty participants had a low (mandatory school) and six a high educational level (university degree). Thirty-two participants reported one or more falls within the last 12 months (12 participants fell once, seven had two falls, 13 fell three times and more in the previous year). Twenty-eight participants didn't fall in the last year (Table 2). The mean and SD of hip and hand-grip strengths are presented in Table 3.

Table 2 : Patient characteristics of our study participants

Variable	Total group (Mean (SD))	Fallers (Mean (SD))	Non-Fallers (Mean (SD))
Sex F/M	38 /22	21/11	17/11
Age (years)	82.0 (6.6)	83.3 (6.2)	80.4 (6.8)
Weight (kg)	65.9 (12.1)	63.8 (12.7)	68.3 (11.2)
BMI	24.3 (3.4)	24.3 (3.9)	24.3 (2.9)
MMSE	26.0 (3.5)	24.3 (3.6)	27.8 (2.1)
FES-I	24.5 (7.1)	26.5 (7.0)	22.2 (6.5)

Abbreviations: BMI, body mass index; MMSE, Mini Mental State Examination; FES-I, falls efficacy scale international; F, female; M, male.

The only hip muscle strength test with an AUC> 0.7 based on the external criterion was the ABD MVIS (0.825, 95% confidence interval (CI) 0.721 to 0.938) (Table 4). The variable importance assessed through the random forest approach (Figure 2) as well showed that the ABD MVIS is the most important variable among the tested muscle groups. Hand-grip strength has an AUC of 0.649 (CI 0.497 to 0.802) and stays below the predefined AUC > 0.7.

Table 3: Mean and standard deviation (SD) for hip strength muscle groups and hand-grip strength presented for the total group as well as the fallers and the non-fallers separately.

	Total group n=60 (Mean (SD))	Fallers n=32 (Mean (SD))	Non-Fallers n=28 (Mean (SD))
ABD MVIS (N/kg)	1.0 (0.5)	0.7 (0.3)	1.2 (0.5)
ABD RFG (N/kg/s)	6.3 (4.5)	4.2 (2.7)	8.8 (5.0)
ADD MVIS(N/kg)	1.1 (0.5)	0.9 (0.4)	1.3 (0.4)
ADD RFG (N/kg/s)	4.4 (3.1)	3.3 (2.3)	5.7 (3.4)
ER MVIS(N/kg)	0.8 (0.3*)	0.7 (0.2**)	0.9 (0.3)
ER RFG (N/kg/s)	3.7 (2.1*)	3.2 (1.9**)	4.3 (2.1)
IR MVIS(N/kg)	1.2 (0.3*)	1.1 (0.3**)	1.2 (0.4)
IR RFG (N/kg/s)	4.2 (2.1*)	3.8 (1.8**)	4.7 (2.4)
EXT MVIS(N/kg)	1.3 (0.5")	1.2 (0.4**)	1.4 (0.6""")
EXT RFG (N/kg/s)	5.9 (3.5")	5.2 (3.7**)	6.7 (3.1""")
FLEX MVIS(N/kg)	1.7 (0.6)	1.4 (0.4)	2.0 (0.6)
FLEX RFG (N/kg/s)	9.9 (5.4)	8.2 (4.9)	11.9 (5.5)
Hand-grip strength (kPa)	52.8 (24.8)	47.1 (21.0)	59.4 (27.4)

* n= 58, " n= 57, ** n=30, "" n=27
MVIC: maximum voluntary isometric strength, RFG: rate of force generation, AUC: area under the curve, CI: confidence interval, ABD: hip abductors, ADD: hip adductors, FLEX: hip flexors, EXT: hip extensors, ER: hip external rotators, IR: hip internal rotators

Table 4: Area under the curve (AUC) for the hip strength muscle groups and hand-grip strength n = 56

Hip muscle group	AUC	95% CI
ABD MVIS	0.825	0.712 to 0.938
ABD RFG	0.787	0.667 to 0.906
ADD MVIS	0.773	0.648 to 0.897
FLEX MVIS	0.755	0.628 to 0.882
ER MVIS	0.741	0.610 to 0.871
ADD RFG	0.706	0.568 to 0.844
FLEX RFG	0.678	0.536 to 0.820
ER RFG	0.651	0.506 to 0.797
Hand-grip strength	0.649	0.497 to 0.802
IR RFG	0.617	0.467 to 0.766
EXT RFG	0.610	0.460 to 0.761
IR MVIS	0.576	0.422 to 0.730
EXT MVIS	0.558	0.404 to 0.712

Abbreviations: MVIS, maximum voluntary isometric strength; RFG, rate of force generation; AUC, area under the curve; CI, confidence interval: ABD, hip abductors strength; ADD, hip adductors strength; FLEX; hip flexors strength; EXT, hip extensors strength; ER, hip external rotators strength; IR, hip internal rotators strength

Figure 2: Mean Decrease Gini sorted decreasingly by variable importance from top to bottom. The two most important variables are the maximum voluntary isometric strength (MVIS) and rate of force generation (RFG) of the hip abductors.

Abbreviations: MVIS; maximum voluntary isometric strength, RFG; rate of force generation; ABD; hip abductors strength, ADD; hip adductors strength, FLEX; hip flexors strength, EXT; hip extensors strength, RI; hip internal rotators strength, ER; hip external rotators strength

DISCUSSION

Our results show that a) hip ABD MVIS has the best ability to distinguish between fallers and non-fallers (based on the fall history) and is the only muscle strength group (AUC of 0.825 (95% CI 0.721 to 0.938)) with an AUC above the predefined level of 0.7, and b) hip frontal plane muscle strength groups (ABD and ADD) distinguish better than hand-grip strength compared to the fall history (hand-grip strength AUC of 0.649 (CI 0.497 to 0.802)).

These results are an important first step to develop a new fall risk assessment tool for older persons. The fall risk of older persons is influenced by multiple factors (38). Therefore it's essential to know the importance of different contributing factors to be able to shape a well-targeted treatment for the persons at risk. Compared to the fall history, hip muscle strength deficits give information about a directly targetable fall risk parameter and contributes to the goal-setting for treatments.

The increasing interest in hip muscle strength as a possible contributing factor to the fall risk of older persons led to some studies in this field (eg (19, 21, 39)). However, none of them gives an indication, which muscle strength variable influences the outcome parameter falls the most. The variable importance of the leading ABD variables in the DGI is followed by the ADD MVIS, but this variable remains, as well as the ABD RFG, below the AUC of 0.7. This particularly supports our hypothesis of the importance of the ABD but also the whole frontal plane muscle strengths. The ADD RFG was ranked clearly lower in the DGI than ADD MVIS. The findings regarding the ABD RFG and ADD RFG seem surprising as a fast recovery step is necessary to regain balance after tripping (40). Therefore, one could assume that the RFG is more important to prevent a fall than MVIS. Literature shows controversial findings according to the importance of RFG (19, 40) and MVIS (21) for the fall risk of older persons. However, a recent article of Morcelli et al (21) allows the presumption to support our findings, although they did not use the same strategy to analyze the data. Morcelli et al (21) assessed hip ABD and ADD MVIS and RFG of older women classified as fallers and non-fallers according to their fall history of the previous 12 months. They identified significant differences between fallers and non-fallers for the late values of RFG of the ABD. Late values of RFG (≥ 90 ms after onset of contraction) are more related to MVIS than to rate of force development (41).

Our second hypothesis regarding the diagnostic accuracy, the comparison between the hip muscle strength and hand grip strength with the same external criterion, was confirmed for ABD and ADD. Hand-grip strength is assumed to be less accurate for fall risk assessment than lower limb strength as it is for example not directly linked to balance recovery (17). Our findings are supported by this study.

In a population of older age, this project was the first which started investigating the strength of all hip muscle groups to identify hip muscle groups above a predefined level of diagnostic accuracy for fall history and consequently their clinical use. The resulting reduction of the total test battery to only one parameter (hip abductor strength), measureable in a short time frame is very important for this age group.

Therefore, this outcome is a perfect starting point. Future research should investigate the value of hip abductor strength for the assessment of the fall risk in older persons. If hip abductor strength shows a clear association with falls a proper evaluation of hip strength deficits besides the assessment of other risk factors should be integrated in the commonly used assessments in order to enable clinicians to design better tailored treatment programs.

The hip abductor measurement with a dynamometer fixed to a custom made frame which is attached to a regular physical therapy treatment table we chose in this study is practicable and feasible with older adults in everyday clinics (20). The time to install and measure participants' hip abductor is 10.58 ± 1.56 (20). Further, our measurement method was proven in our earlier article to have a good reliability (20).

Our investigation of the hip strength muscle groups in hospitalized and outpatients allowed to get information about the full range of applicability of this new measurement tool and thereby is a strong point for a wide range of possible application areas. The reduced test battery, the easily practicable measurement method in daily clinics as well as the wide applicability range further facilitates the investigation and implementation of this easily targetable fall risk parameter.

A point to consider in future investigations of ABD MVIS as a fall risk assessment tool is the fall prevalence in the sample under study. One part of our participants was recruited in a geriatric hospital known for its fall risk prevention interventions. This lead to our higher fall prevalence (53%) as the worldwide reported fall rate of the WHO for older persons of 28 to 35% (5). This higher prevalence only slightly influences the results of the AUC and DGI but can possibly lead to a small overestimation of the diagnostic accuracy of a measurement instrument.

Our sample size was calculated prospectively to achieve a power of 80%, nonetheless, 60 participants stay a rather small sample size for this heterogeneous population. Therefore, to allow a wide clinical use hip ABD MVIS should further be investigated in a broader setting of older persons to gather insight in its diagnostic accuracy for the daily clinical use as a fall risk assessment tool. Lastly, we assessed our participants MVIS and RFG in the same trial and only found a slightly weaker ability to distinguish between fallers and non-fallers for the ABD RFG compared to the ABD MVIS. Due to the complexity of this task for older persons, this combined measurement of MVIS and RFG could have influenced the study results and should

be investigated in two different trials to allow a stricter differentiation of the results for MVIS and RFG.

CONCLUSION

Hip ABD seemed to be an important hip muscle group to distinguish between older adult fallers and non-fallers compared with the chosen external criterion, history taking of falling. Hand-grip strength has a lower ability to distinguish compared to fall history than ABD and ADD in our population. Future studies should focus on the diagnostic accuracy of ABD MVIS and a specific hip ABD muscle training to decrease fall risk.

ACKNOWLEDGEMENT

We kindly thank all our study participants for their participation as well as all physical therapists who assisted us during patient recruitment.

CONFLICT OF INTEREST

none declared

DISCLOSURE

The abstract of this paper was presented the 19th of October 2017 as a poster presentation with interim findings at the "1. cooperation congress of Reha Schweiz and Physioswiss" in Davos, Switzerland.

REFERENCES

1. Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC geriatrics*. 2011;11:33.
2. Tinetti ME, Williams CS. Falls, injuries due to falls, and the risk of admission to a nursing home. *The New England journal of medicine*. 1997;337(18):1279-84.
3. Muller M, Maier AB, Smulders YM. [High blood pressure and mortality in the elderly: what does gait speed tell?]. *Nederlands tijdschrift voor geneeskunde*. 2013;157(7):A5801.
4. Liu SW, Obermeyer Z, Chang Y, Shankar KN. Frequency of ED revisits and death among older adults after a fall. *The American journal of emergency medicine*. 2015;33(8):1012-8.
5. WHO World Health Organization report/ WHO Global report on falls prevention in older age. 2007. ISBN 978 92 4 156353 6, 2007.
6. Swiss Federal Statistical Office. Schweizerische Gesundheitsbefragung 2012; Die funktionale Gesundheit von älteren Menschen in Privathaushalten. 2014.
7. Swiss Federal Statistical Office. Gesundheit von Betagten in Alters- und Pflegeheimen Erhebung zum Gesundheitszustand von betagten Personen in Institutionen (2008/09). 2012.
8. Muir SW, Berg K, Chesworth BM, Klar N, Speechley M. Modifiable Risk Factors Identify People Who Transition from Non-fallers to Fallers in Community-Dwelling Older Adults: A Prospective Study. *Physiother Can*. 2010;62(4):358-67.
9. Batista FS, Gomes GA, D'Elboux MJ, Cintra FA, Neri AL, Guariento ME, et al. Relationship between lower-limb muscle strength and functional independence among elderly people according to frailty criteria: a cross-sectional study. *Sao Paulo medical journal = Revista paulista de medicina*. 2014;132(5):282-9.
10. Hanewinkel R, van Oijen M, Ikram MA, van Doorn PA. The epidemiology and risk factors of chronic polyneuropathy. *European journal of epidemiology*. 2016;31(1):5-20.
11. Hoffman EM, Staff NP, Robb JM, St Sauver JL, Dyck PJ, Klein CJ. Impairments and comorbidities of polyneuropathy revealed by population-based analyses. *Neurology*. 2015;84(16):1644-51.
12. Allet L, Kim H, Ashton-Miller J, De Mott T, Richardson JK. Frontal plane hip and ankle sensorimotor function, not age, predicts unipedal stance time. *Muscle & nerve*. 2012;45(4):578-85.
13. Liaw MY, Chen CL, Pei YC, Leong CP, Lau YC. Comparison of the static and dynamic balance performance in young, middle-aged, and elderly healthy people. *Chang Gung medical journal*. 2009;32(3):297-304.
14. Maki BE, Holliday PJ, Topper AK. A prospective study of postural balance and risk of falling in an ambulatory and independent elderly population. *Journal of gerontology*. 1994;49(2):M72-84.
15. Rogers MW, Hedman LD, Johnson ME, Cain TD, Hanke TA. Lateral stability during forward-induced stepping for dynamic balance recovery in young and older adults. *J Gerontol A Biol Sci Med Sci*. 2001;56(9):M589-94.
16. Tinetti ME, Kumar C. The patient who falls: "It's always a trade-off". *Jama*. 2010;303(3):258-66.
17. Pijnappels M, van der Burg PJ, Reeves ND, van Dieen JH. Identification of elderly fallers by muscle strength measures. *Eur J Appl Physiol*. 2008;102(5):585-92.
18. Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, Knol DL. *Measurement in medicine, a practical guide*: Cambridge University Press; 2011. 338 p.
19. Palmer TB, Thiele RM, Williams KB, Adams BM, Akehi K, Smith DB, et al. The identification of fall history using maximal and rapid isometric torque characteristics of the hip extensors in healthy, recreationally active elderly females: a preliminary investigation. *Aging Clin Exp Res*. 2015;27(4):431-8.
20. Gafner S, Bastiaenen CHG, Terrier P, Punt I, Ferrari S, Gold G, et al. Evaluation of hip abductor and adductor strength in the elderly: a reliability study. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity*. 2017;14:5.
21. Morcelli MH, Rossi DM, Karuka AH, Crozara LF, Hallal CZ, Marques NR, et al. Peak torque, reaction time, and rate of torque development of hip abductors and adductors of older women. *Physiotherapy theory and practice*. 2016;32(1):45-52.
22. Mendis MD, Wilson SJ, Hayes DA, Watts MC, Hides JA. Hip flexor muscle size, strength and recruitment pattern in patients with acetabular labral tears compared to healthy controls. *Manual therapy*. 2014;19(5):405-10.
23. Kollock RO, Jr., Onate JA, Van Lunen B. The reliability of portable fixed dynamometry during hip and knee strength assessments. *Journal of athletic training*. 2010;45(4):349-56.
24. Thorborg K, Petersen J, Magnusson SP, Holmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. *Scandinavian journal of medicine & science in sports*. 2010;20(3):493-501.
25. Krause DA, Schlagel SJ, Stember BM, Zoetewey JE, Hollman JH. Influence of lever arm and stabilization on measures of hip abduction and adduction torque obtained by hand-held dynamometry. *Archives of physical medicine and rehabilitation*. 2007;88(1):37-42.
26. Desrosiers J, Bravo G, Hebert R, Dutil E. Normative data for grip strength of elderly men and women. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association*. 1995;49(7):637-44.
27. Burgi F, Niederer I, Schindler C, Bodenmann P, Marques-Vidal P, Kriemler S, et al. Effect of a lifestyle intervention on adiposity and fitness in socially disadvantaged subgroups of preschoolers: a cluster-randomized trial (Ballabeina). *Preventive medicine*. 2012;54(5):335-40.
28. Young WR, Mark Williams A. How fear of falling can increase fall-risk in older adults: applying psychological theory to practical observations. *Gait & posture*. 2015;41(1):7-12.

29. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). *Age and ageing*. 2005;34(6):614-9.
30. Muir SW, Gopaul K, Montero Odasso MM. The role of cognitive impairment in fall risk among older adults: a systematic review and meta-analysis. *Age and ageing*. 2012;41(3):299-308.
31. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975;12(3):189-98.
32. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *Journal of pediatric psychology*. 2014;39(2):204-21.
33. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of biomedical informatics*. 2014;48:193-204.
34. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
35. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine*. 2000;45(1-2):23-41.
36. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32.
37. Liaw A, Wiener M. Classification and regression by randomForest. . *R News* 2. 2002;3:18-22
38. Tromp AM, Pluijm SM, Smit JH, Deeg DJ, Bouter LM, Lips P. Fall-risk screening test: a prospective study on predictors for falls in community-dwelling elderly. *Journal of clinical epidemiology*. 2001;54(8):837-44.
39. Morcelli MH, Crozara LF, Rossi DM, LaRoche DP, Marques NR, Hallal CZ, et al. Hip muscles strength and activation in older fallers and non-fallers. *Isokinetics and Exercise Science*. 2014;22:191-6.
40. Bento PC, Pereira G, Ugrinowitsch C, Rodacki AL. Peak torque and rate of torque development in elderly with and without fall history. *Clinical biomechanics (Bristol, Avon)*. 2010;25(5):450-4.
41. Andersen LL, Aagaard P. Influence of maximal muscle strength and intrinsic muscle contractile properties on contractile rate of force development. *European journal of applied physiology*. 2006;96(1):46-52.

CHAPTER 5

The role of hip abductor strength in identifying older persons at risk of falls: A diagnostic accuracy study

Simone C. Gafner, Caroline H. G. Bastiaenen, Serge Ferrari, Gabriel Gold, Andrea Trombetti, Philippe Terrier, Roger Hilfiker, Lara Allet

Clinical interventions in aging. 2020;15:645-54.

DOI: 10.2147/CIA.S246998

ABSTRACT

Background/Objectives:

Early detection of fall risk in persons older than 65 is of clinical relevance, but the diagnostic accuracy of currently used functional tests (e.g. Short Physical Performance Battery [SPPB] and Timed Up and Go test [TUG]) to assess older persons' fall risks remains moderate.

Recent literature highlights the importance of strong hip abductors to prevent falls. We thus aimed to assess the diagnostic accuracy of hip abductor strength measures to assess older persons' fall risks.

Methods:

Hip abductor maximum voluntary isometric strength (ABD MVIS), rate of force generation (ABD RFG), and the SPPB and TUG functional fall risk assessments were assessed in 60 persons aged over 65 years (82.0 ± 6.1 years). The diagnostic accuracy (area under the curve [AUC], sensitivity [sens], specificity [spec], positive predictive value [PPV], negative predictive value [NPV], and positive and negative likelihood ratios [LR⁺, LR⁻]) was evaluated at a clinically important 90% sensitivity level. Cut-off values for clinical use were calculated.

Results:

In our population, hip ABD MVIS (AUC 0.8, sens 90.6%, spec 57.1%, PPV 70.7%, NPV 84.2%, LR⁺ 2.1, LR⁻ 0.2, and cut-off value ≤ 1.1 N/kg) and hip ABD RFG (AUC 0.8, sens 90.6%, spec 46.4%, PPV 65.9%, NPV 81.3%, LR⁺ 1.7, LR⁻ 0.2, and cut-off ≤ 8.47 N/kg/s) show diagnostic accuracy comparable to other fall risk assessments (SPPB and TUG) and a high net sensitivity when used in a test battery.

Conclusion:

Hip ABD MVIS or RFG shows good diagnostic accuracy to differentiate between older fallers and nonfallers compared to the chosen external criterion history of falls. The high net sensitivity when hip ABD MVIS or RFG are combined with currently used fall risk assessments shows promise in contributing value to a test battery and should be investigated further in longitudinal studies.

Keywords: measurement study, muscle strength, hip, functional performance, accidental falls, aged

INTRODUCTION

The World Health Organization (WHO) reports a yearly fall rate of 28 to 35% in persons older than 65 (1). For persons living in long term care institutions, percentages are even higher and lay between 30 to 50% (1).

Various functional tests are currently used to assess older persons' fall risks (e.g. Timed Up and Go Test [TUG] and Short Physical Performance Battery [SPPB]). However, the diagnostic accuracy (e.g. sensitivity and specificity) of these functional tests remains moderate (2, 3). In addition, poor performance in a functional test indicates a fall risk but requires further tests to identify the underlying causes. It is thus of utmost importance to develop and use valid diagnostic fall risk assessment tools that can be used in everyday clinical practice and that indicate parameters that need to be targeted.

A combination of various factors often leads to a fall incident (e.g. muscle weakness, gait and balance problems, poor footwear, or poor lighting) (4), which can cause serious consequences such as fractures, loss of independence (5), and increased mortality (6). Muscle strength has already been shown to be important in preventing falls (4). Recent literature further underlines the significance of hip abductor muscle strength to maintain mediolateral balance control (7-9), which is essential to avoid lateral and posterolateral falls (10). This is of clinical relevance because posterolateral falls have high hip injury potential and lead to fractures (11-13). In addition, it was shown that hip abductor muscle contractions at the moment of impact after a fall could substantially reduce stress at the femoral neck and likewise the risk of a femoral fracture (14). Therefore, hip abductor strength could be a promising contributing factor for the assessment of fall risk and establishment of cut-off values, and diagnostic accuracy might improve the clinical decision-making process regarding individuals' fall risks.

Our previous work confirmed the importance of hip abductor muscle strength and showed that the strength of the hip abductor muscle groups can, among all hip muscle groups (flexors, extensors, internal and external rotators and adductors), best distinguish between older fallers and nonfallers (15).

Therefore, the current study first aimed to investigate diagnostic accuracy (area under the curve [AUC], sensitivity [sens], specificity [spec], positive predictive value [PPV], negative predictive value [NPV], and positive and negative likelihood ratio [LR⁺, LR⁻]) of hip abductor strength (ABD maximum voluntary isometric strength [MVIS] and rate of force generation [RFG]) to distinguish between older fallers and nonfallers compared to a history of falling as the predefined reference standard. The second goal was to identify clinically relevant hip abductor strength (MVIS and RFG) cut-off values based on diagnostic accuracy outcomes to distinguish between older

fallers and nonfallers evaluated by fall history as the same reference standard. The third aim was to compare the diagnostic accuracy of hip abductor strength with the diagnostic accuracy of currently used fall risk assessments (TUG and SPPB) in our population, as well as with published articles. The final goal was to calculate the net sensitivity and specificity of our population for parallel testing of ABD MVIS or ABD RFG with the mentioned currently used TUG and SPPB fall risk assessments.

We tested three hypotheses. First, hip abductor strength has a good diagnostic accuracy (≥ 0.9) for clinical use to distinguish between older fallers and nonfallers. Second, the diagnostic accuracy of hip abductor strength shows values comparable to the diagnostic accuracy of currently used fall risk assessments (TUG and SPPB). Third, hip abductor strength (ABD MVIS and RFG), as a not yet systematically assessed parameter in the construct of fall risk, shows high accuracy in recognizing fallers in parallel testing with TUG and SPPB.

METHODS

We performed a validity study to evaluate the diagnostic accuracy of hip ABD MVIS and ABD RFG for the fall risk evaluation of persons' older than 65. The Standards for Reporting of Diagnostic Studies (STARD) (16) were followed (Figure 1 shows the STARD flow chart to report participants' flow through the study). Participants were simultaneously recruited from the Geriatric Department of the University Hospitals of Geneva and from an ambulatory setting. They were included when they were ≥ 65 years, had a body mass index (BMI) between 17 and 35 and were able to walk household distances. We excluded participants in case of a positive history or evidence of any significant central nervous system dysfunction, neuromuscular disorders except a distal symmetric peripheral neuropathy, evidence of vestibular dysfunction, moderate or severe dementia (Minimal Mental State Exam (MMSE) < 18), a fracture of the lower limb and/ or joint replacement within the previous year and any lower limb or back pain that adversely affects the strength tests (15).

This study is an extended evaluation and a next step to the results we presented in our previous article in which, as a first step, we evaluated which of the hip muscle groups is the most important to differentiate between older fallers and nonfallers (15). In the current study, we examine the diagnostic accuracy of hip abductor muscle strength to decide whether this parameter should be integrated in a test battery to assess older persons' fall risks. The diagnostic accuracy of a test is expressed by the area under the curve [AUC], sensitivity [sens], specificity [spec], positive predictive value [PPV], negative predictive value [NPV], and positive and negative likelihood ratio [LR+, LR-])

Figure 1: STARD diagram of participants flow through the study

The area under the receiver operating curves is the AUC, the higher the value of the AUC the better a test discriminates between two groups (17). Sensitivity and specificity reflect the intrinsic quality of a test and show the ability of a measurement instrument to detect true positives or true negatives respectively (18). The likelihood ratio is calculated based on sensitivity and specificity (18). The positive likelihood ratio is a value for how much more likely a positive test result is in people with the disease than in people without the disease (18). How much less likely a negative test result is in people with the disease than in people without the disease is expressed by the negative likelihood ratio (18). The larger the value above one (for LR+), respectively

the smaller the value below one (for LR-), the more valuable is the test result (18). Positive and negative predictive values are the probability of the disease in a patient with a positive test result or the probability of not having the disease when the test result is negative respectively (17).

The study was conducted according to the Declaration of Helsinki and the local ethics committee (commission cantonal d'éthique de la recherche Geneva, CCER) approved the study protocol (14-235). Participants received verbal and written information about the study content, and they signed informed consent forms prior to the first tests.

Procedure

All of the included patients' demographic data (e.g. age, sex, and BMI) were extracted from the clinical database. The reference standard fall history was collected by verbal questioning.

Thereafter, trained physical therapists (blinded to the participants' fall histories) assessed the Falls Efficacy Scale International (FES-I), (19) Mini Mental State Examination (MMSE) (20), the SPPB (21) and TUG (22) functional fall risk assessments, as well as participants' hip abductor strength (MVIS and RFG) following our study protocol's procedure.

INDEX MEASURE: HIP ABDUCTOR STRENGTH

Hip abductor strength was measured in Newtons (N) using an analog dynamometer (Sensix, Poitiers, France) with a range from 0 to 667 N and a precision of 0.002 N (23). The Delsys System (Trigno sensor, Delsys, Inc., Boston, MA) coupled with the calibrated dynamometer digitalized the output (3.3 V) with a sampling rate of 1926 Hz and a 16-bit resolution (23).

For hip abductor strength (ABD) tests, participants lay on their sides and performed hip ABD with the upper leg in a 10° abducted position (23). The lower leg was at 45° of knee and hip flexion for stabilization. Participants were asked to keep the knee of their tested leg extended with the hip in a neutral or slightly extended position (23). For the test, the examiner manually supported the leg's starting position 1 cm from the dynamometer that was positioned 5 cm proximal to the malleolus externus (23, 24).

To measure MVIS and RFG, participants were asked to push as hard and fast as possible against the dynamometer (25), hold the pressure for 3s, and relax (26). To avoid bias through dynamometer destabilization or the testing person's strength,

the dynamometer was fixed to a metallic frame at the appropriate height for the described 10° abduction starting position.

Participants repeated the test 3 times per leg with a break of 1 min between trials. In accordance with our pilot tests, the mean of the 6 trials was retained for further analyses. Details of the measurement procedure and information about the feasibility and reliability of the test are described elsewhere (15, 23).

FUNCTIONAL FALL RISK ASSESSMENTS:

Short Physical Performance Battery (SPPB): The SPPB estimates the function of the lower limbs. It is divided in 3 parts: the evaluation of standing balance, walking speed, and lower limb strength with a 5 sit to stand test (21, 27). The final score ranges from 0 to 12 points and is calculated by summarizing the scores of the 3 parts.

Timed Up and Go Test: For the TUG, participants were asked to get up from a chair with armrests, walk 3 m at usual gait speed, perform a turn, and return to the chair to sit down again. Participants were allowed a test trial. Thereafter, the time needed to complete the task once was recorded in seconds (28).

REFERENCE STANDARD: FALL HISTORY

To classify our study participants into fallers and nonfallers, a history of falls during the previous 12 months was chosen as a reference standard. A faller was defined as a participant who reported one or more falls within the previous year (29). Participants who did not report a fall were classified as nonfallers (29). Participants were informed that a fall was defined as an event resulting in a person inadvertently coming to rest on the ground, floor, or other lower level (1). We didn't include falls resulting from unavoidable environmental hazards, such as a chair collapsing (22).

Sample size

We expected to find high sensitivity levels comparable with currently used fall risk assessment tools (sensitivity between 0.6 and 0.9). To test the null hypothesis, H_0 : sensitivity = 0.6 (H_1 : sensitivity $\neq$ 0.6) with an expected sensitivity of 0.9, a minimum sample of $n = 19$ persons with the disease (fallers) and a total sample of $n = 48$ is appropriate to achieve a <5% significance level and 80% power (30). To compensate for an expected 20% dropout rate, we aimed to include 60 participants.

Statistics

Statistical analyses were performed with Stata version 14.1 (College Station, Texas 77845, USA).

Data Processing

After downloading raw data to a computer, Matlab (Mathworks, Natick, MA, USA) was used to process the force data. A low-pass filter (75 ms moving average) suppressed high-frequency noise. The MVIS was defined as the highest force generated by the subject within 3s from the beginning of the test. The RFG was evaluated within a 50ms time window after 10% of the MVIS was reached (23). MVIS and RFG were normalized to body mass (15).

Main analyses

The characteristics of the study sample were described with mean and standard deviations. Area under the curves were calculated from logistic regressions with their corresponding 95% confidence intervals (95% CI).

To achieve a high probability to detect as many persons at risk of falls as possible, high test sensitivity is important (i.e. a test's ability to detect affected persons). A way to increase the "certainty" of detecting at-risk persons is to perform two different tests simultaneously (parallel testing), which leads to a gain in sensitivity (net sensitivity).

To obtain a high number of true positives (and consequently a lower number of false negatives), we decided to evaluate the diagnostic accuracy of our index measure closest above the clinically important 90% sensitivity level for discrimination. Therefore, sensitivity and specificity (with corresponding 95% confidence intervals) were calculated based on cut-offs closest to the value that yielded in a sensitivity of ≥ 0.9 . Positive and negative predictive values, as well as likelihood ratios for positive (LR+) and negative (LR-) tests were calculated (with corresponding 95% confidence intervals) accordingly.

For the calculations of the net sensitivity and net specificity, we additionally present the calculations with the Youden Index, Youden's J ($J = \text{Sensitivity} + \text{Specificity} - 1$) with the corresponding cut-offs. We further analyzed and presented indices for discrimination and calibration of our logistic regression models.

Finally, we also calculated calibration plots of the logistic regression model with the Hosmer-Lemeshow test, with a significance level (type-I error) of 5%. We plotted our observed values for falls against the predicted probabilities of falls for groups defined by ranges of 20% for the predicted risk, as well as their slopes and intercepts (31).

RESULTS

We evaluated the diagnostic accuracy of hip abductor MVIS and RFG in 60 participants (mean age 82 years, SD 6.61 years, 28 older nonfallers and 32 older fallers). The descriptive statistics for age, sex, BMI, FES-I, MMSE, SPPB, and TUG are presented in Table 1.

Table 1: Characteristics of our study participants

Variable	Total group mean (SD)	Fallers mean (SD)	Nonfallers mean (SD)
Sex (F/M)	38/22	21/11	17/11
Age (years)	81.2 (6.6)	83.3 (6.2)	80.4 (6.8)
Weight (kg)	65.9 (12.1)	63.8 (12.7)	68.3 (11.2)
BMI (kg/m ²)	24.3 (3.4)	24.3 (3.9)	24.3 (2.9)
Hospitalized (%)	71.7	96.9	42.9
MMSE	26.0 (3.5)	24.3 (3.6)	27.8 (2.1)
FES-I	24.5 (7.1)	26.5 (7.0)	22.2 (6.5)
SPPB	8.2 (3.1)	6.3 (2.3)	10.3 (2.4)
TUG (s)	15.9 (9.0)	20.0 (9.4)	11.3 (5.8)

Abbreviations: BMI, body mass index; MMSE, mini mental state examination; FES-I, falls efficacy scale international; SPPB, short physical performance battery; TUG, timed up and go test; s, seconds

The diagnostic accuracy (AUC, sens, spec, PPV, NPV, and LR+ and LR-) to discriminate between older fallers and nonfallers using hip abductor MVIS and RFG as well as the corresponding cut-off values of our sample are presented in Table 2.

The diagnostic accuracy (AUC, sens, spec, PPV, NPV, and LR+ and LR-) to discriminate between older fallers and nonfallers using hip abductor MVIS and RFG as well as the corresponding cut-off values of our sample are presented in Table 2. The diagnostic accuracy to distinguish between older fallers and nonfallers of ABD MVIS, ABD RFG, TUG, and SPPB shows comparable values (values range between the following: Sens 90.6-93.9%, Spec 46.4-71.5%, PPV 65.9-78.4%, NPV 81.3.-88.2%, LR+ 1.7-3.2, LR- 0.1-0.2). However, hip ABD MVIS shows a slightly higher diagnostic accuracy (AUC 0.8 [95% CI, 0.7-0.9], Sens 90.6% [75.0-98.0], Spec 57.1% [37.2-75.5], PPV 70.7% [51.8-92.5], NPV 84.2% [62.3-92.5], LR+ 2.1 [1.4-3.3], LR- 0.2 [0.1-0.5]) than hip ABD RFG (AUC 0.8 [95% CI, 0.7-0.9], Sens 90.6% [75.0-98.0], Spec 46.4%[27.5-66.1], PPV 65.9%[45.8-90.8], NPV 81.3% [57.3-90.7], LR+ 1.7 [1.2-2.4], LR- 0.2 [0.1-0.6]) (Table 2). The cut-off values to distinguish between fallers and nonfallers are ≤ 1.1 N/kg for ABD MVIS and ≤ 8.5 N/kg/s for ABD RFG.

In Table 3, the cross-tabulation of the index tests (hip ABD MVIS or RFG) are presented against the results of the reference standard. This allows to identify the true positive, true negative, false positive and false negative cases.

Table 2: The diagnostic accuracy to discriminate between older fallers and nonfallers of hip abductor strength (MVIS and RFG) and the performance tests SPPB and TUG, as well as their corresponding cut-off values evaluated by the closest sensitivity value > 90%.

Variable	AUC (95% CI)	Sens % (95% CI)	Spec % (95% CI)	PPV % (95%CI)	NPV % (95% CI)	LR+ (95% CI)	LR- (95% CI)	Cut-off
ABD MVIS (N/ kg)	0.8 (0.7, 0.9)	90.6 (75.0, 98.0)	57.1 (37.2, 75.5)	70.7 (51.8, 92.5)	84.2 (62.3, 92.5)	2.1 (1.4, 3.3)	0.2 (0.1, 0.5)	≤ 1.1
ABD RFG (N/kg/s)	0.8 (0.7, 0.9)	90.6 (75.0, 98.0)	46.4 (27.5, 66.1)	65.9 (45.8, 90.8)	81.3 (57.3, 90.7)	1.7 (1.2, 2.4)	0.2 (0.1, 0.6)	≤ 8.5
SPPB (12)	0.9 (0.8, 1.0)	90.6 (75.0, 98.0)	71.4 (51.3, 86.8)	78.4 (60.5, 94.9)	87.0 (67.4, 94.6)	3.2 (1.7, 5.8)	0.1 (0.0, 0.4)	≤ 9.0
TUG (sec)	0.8 (0.7, 0.9)	93.8 (79.2, 99.2)	53.6 (33.9, 72.5)	69.8 (53.9, 82.8)	88.2 (63.6, 98.5)	2.0 (1.3, 3.0)	0.1 (0.0, 0.5)	≥ 10.0

Abbreviations: AUC: area under the curve; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio; 95% CI ; 95% confidence interval; ABD MVIS, hip abductor maximum voluntary isometric strength; ABD RFG, hip abductor rate of force generation; SPPB, short physical performance battery; TUG, timed up and go test; sec, seconds; N, Newton; s, seconds

Table 3: Cross tabulation of the index measure (left; ABD MVIS, right; ABD RFG) against the reference standard history of falling

Hip ABD MVIS ≤ 1.1 N/kg	Fall history		Hip ABD RFG ≤ 8.47 N/kg/s		Fall history	
	Positive	Negative	Positive	Negative	Positive	Negative
Positive	29	12	41	29	15	44
Negative	3	16	19	3	13	16
Total	32	28	60	32	28	60

Abbreviations: ABD MVIS, hip abductor maximum voluntary isometric strength; ABD RFG, hip abductor rate of force generation

Our ABD MVIS and RFG model, as well as SPPB and TUG, showed good calibration (small discrepancy between predicted probability of falls on the x-axis and observed proportion of falls on the y-axis, all P values > 0.05, Figure 2). This means that for example for the hip ABD MVIS values (Figure 2a) for the participants we predicted about a 10% risk of falls (x-axis) we really observed a risk of falls of about 10% (y-axis). A perfect prediction lays on the 45° line. The net sensitivities and specificities for the parallel testing of the currently used TUG and SPPB fall risk assessment tools with hip ABD MVIS and RFG of our population are presented in Table 4 (once for sensitivity ≥90% and once for the sensitivity evaluated with the Youden Index). Table 4 shows that performing 2 tests simultaneously (ABD MVIS or ABD RFG combined with either TUG or SPPB) leads to a very high net sensitivity (between 96.5 and 99.4) in our study sample.

Figure 2: The calibration plots for the abductor maximum voluntary isometric strength (ABD MVIS), hip abductor rate of force generation (ABD RFG), the short physical performance battery (SPPB) and the timed up and got test (TUG) show good calibration (small discrepancy between predicted probability of falls and observed proportion of falls, all p-values > 0.05).

Table 4: Calculation of net sensitivity for parallel testing of one of the currently used fall risk assessment tools (TUG or SPPB) with the index measure (ABD MVIS or ABD RFG) of our population.

Current fall risk assessment tool	Sens (%)	Spec (%)	Index measure	Sens (%)	Spec (%)	Net Sens for parallel testing (%)	Net Spec for parallel testing (%)
Sensitivity ≥90							
TUG	93.8	53.6	ABD MVIS	90.6	57.1	99.4	30.6
TUG	93.8	53.6	ABD RFG	90.6	46.4	99.4	24.9
SPPB	90.6	71.4	ABD MVIS	90.6	57.1	99.1	40.8
SPPB	90.6	71.4	ABD RFG	90.6	46.4	99.1	33.2
Sensitivity Youden Index							
TUG	84.4	75.0	ABD MVIS	78.1	82.1	96.6	61.6
TUG	84.4	75.0	ABD RFG	75.0	71.4	96.1	53.6
SPPB	87.5	78.6	ABD MVIS	78.1	82.1	97.3	64.5
SPPB	87.5	78.6	ABD RFG	75.0	71.4	96.9	56.1

Abbreviations: TUG, timed up and go test; SPPB, short physical performance battery; ABD MVIS, hip abductor maximum voluntary isometric strength; ABD RFG, hip abductor rate of force generation

DISCUSSION

We hypothesized the following; first, for clinical use, hip abductor strength has good diagnostic accuracy to distinguish between older fallers and nonfallers. Second, the diagnostic accuracy of hip abductor strength is comparable with the diagnostic accuracy of currently used fall risk assessments (TUG and SPPB). Third, hip abductor strength (ABD MVIS and RFG) shows good values by which to recognize fallers in parallel testing with TUG and/or SPPB.

Our results support our hypotheses. In our population, hip ABD strength shows good diagnostic accuracy to distinguish between older fallers and nonfallers, comparable to values of other functional fall risk and/or mobility assessment tools. Hip ABD MVIS (AUC 0.8 [95% CI, 0.7-0.9], cut-off value ≤ 1.1 N/kg) shows a slightly higher diagnostic accuracy than hip ABD RFG (AUC 0.8 [95% CI, 0.7-0.9], cut-off ≤ 8.5 N/kg/s) (Table 2). ABD MVIS's slightly better diagnostic accuracy compared to ABD RFG seems rather unexpected as clinically one can assume that RFG is more important in fall prevention since a fast recovery step after tripping is necessary to regain balance and prevent falls (32). However, this might be because our participants' ABD MVIS and RFG were evaluated within the same trial. For older persons, our combined instructions to push as hard and as fast as possible to

evaluate ABD MVIS and ABD RFG in the same trial was very complex. Therefore, for a clear differentiation of the 2 variables (MVIS and RFG), we recommend a separation of the trials to evaluate ABD MVIS and ABD RFG in future studies.

For clinical use, a quick and easy assessment with good diagnostic accuracy is essential. So far, various studies have assessed the diagnostic accuracy of the TUG or the SPPB to assess older persons' fall risks, 2 tests that are easily applicable in a relatively short timeframe (2, 33). Within our population, the ABD MVIS, ABD RFG, TUG, and SPPB showed comparable diagnostic accuracy (Table 2). This might be surprising as the hip ABD MVIS and ABD RFG evaluate only one specific dimension of the 3 often-named fall-influencing parameters (i.e. strength, balance, and gait characteristics) (34). Nevertheless, the influence of the hip abductors on older persons' decreased mediolateral balance control (7-9) and the femoral-neck-stress-reducing hip abductor muscle contraction at the moment of impact after falling (14) gives an indication of the importance of this hip muscle group in relation to falls. Therefore, evaluating and understanding the construct of older persons' fall risks might benefit from the additional assessment of hip abductor strength as a not yet systematically assessed parameter to currently used fall risk assessments evaluating complementary aspects of falls.

In Barry et al. (2) systematic review, including retrospective and prospective studies of community-dwelling persons, the TUG showed an overall AUC of 0.57 (95% CI, 0.54-0.59) to discriminate between fallers and nonfallers. In another study, the AUC of the TUG to differentiate between persons who fell 6 months prior to hospitalization and persons who did not fall was 0.58 (34). The AUC for prospectively assessed in-hospital and in-hospital injurious falls ranged between 0.64 and 0.67 (34). The AUC of the TUG in our study was 0.8 (95% CI, 0.7-0.9), which is somewhat higher than the AUC in the aforementioned articles.

In Lauretani et al. (3) the SPPB showed an AUC of 0.7 (95% CI, 0.6-0.7) to differentiate between persons who fell in the previous 12 months and nonfallers in an outpatient population (3). The AUC of the SPPB to identify persons who fell in the 6 months prior to hospitalization was 0.57, and in the same study, it was between 0.68 and 0.72 for the prospective assessment of in-hospital and in-hospital injurious falls (34). For our population, we calculated an AUC of 0.8 (95% CI, 0.7-0.9), which is again higher than the previously stated values. One explanation for the higher AUC in our study might be the rather heterogeneous study population because we recruited in- and outpatients; it is likely that the different patient characteristics (e.g. results of the functional performance tests and hospitalization rate) could have led to a slight over- or underestimation of the discriminating ability of hip abductor strength. However, to the best of our knowledge, this study is one of the first that has calculated the diagnostic accuracy of hip abductor strength to differentiate between

older persons at risk of falls and nonfallers. The inclusion of hospital inpatients and outpatients in a first step allowed the analysis of the variable hip ABD on its full applicability range. Future studies should thus confirm our result using longitudinal designs with a more homogenous sample.

A high sensitivity and specificity is of utmost clinical importance when assessing older persons' fall risks. Regarding the deleterious adverse effects a fall can have for older persons, we chose to determine the diagnostic accuracy by using a high preselected sensitivity of $\geq 90\%$ instead of the often-used Youden Index. Our chosen approach to calculating the cut-off values implies accepting a small risk of treating too many rather than not enough older persons. For every test used to differentiate between older persons at risk or not at risk of falls, a very wide range of cut-off values is published, indicating that more than one fall risk test is necessary to classify a person into the categories "at risk" or "not at risk." To increase the certainty of detecting as many persons at risk of falls as possible, net sensitivity and specificity calculations should be performed. Parallel testing (performing 2 different tests simultaneously) increases the ability to detect persons at risk of falling (gain in net sensitivity but decrease in net specificity) and allows identification of the most accurate test battery to assess older persons at fall risk. Therefore, we calculated the net sensitivity and specificity when using one of the commonly used fall risk assessment tools (SPPB and TUG) in parallel testing with hip ABD MVIS or ABD RFG in our study population. To make rendering comparisons with other studies easier, we also presented the calculation of the net sensitivity with the sensitivity evaluated by the Youden Index. The calculated net sensitivities in our study (between 96.5 and 99.4) are high and show a good ability of the hip ABD MVIS and RFG combined with the TUG and SPPB to detect persons at risk of falls. On the other hand, the net specificities are lower, indicating that the combination of hip ABD MVIS or RFG with the TUG or SPPB is not useful to detect nonfallers.

Hip abductor strength is easily and reliably assessable (23). The inclusion of hospital inpatients and outpatients in this study allowed the analysis of the variable hip ABD on its full applicability range. Our chosen high sensitivity ($\geq 90\%$) is a further strength of this study because our choice was to reduce the risk of excluding persons at risk of falls.

A limit of the present study is that the fall history data were not collected in a prospective way, in a rather heterogeneous sample of older participants, which might lead to a slight under- or overestimation of the diagnostic accuracy. It might be that a recall bias influenced the reported number of falls, which might have led to a misclassification. Our strength data are normalized to body mass but we didn't take into account the specific weight and length of the participants' leg. This might have influenced the cut-off points of the ABD MVIS and RFG respectively.

The above-mentioned results suggest that a proper hip abductor strength assessment can be used as a contribution to a test battery to assess patients' fall risks and indicate a possible weakness.

CONCLUSION

Hip abductor strength (MVIS and RFG) shows good diagnostic accuracy to differentiate between older fallers and nonfallers. The high net sensitivity when hip ABD strength measurements (MVIS and RFG) are combined with currently used fall risk assessments shows its promising contributing value to a test battery. Hip abductor strength is an easily measurable parameter and should on a daily clinically basis be considered as an important influencing factor of fall risk. However, further research is needed. Future studies should find pragmatic solutions to reliably and validly assess hip abductor strength in a standard clinical setting and assess the validity of hip abductor strength to predict fall risk in longitudinal studies.

ACKNOWLEDGMENTS

The authors thank all study participants for their participation, as well as all physical therapists who assisted us during this study.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

SPONSOR'S ROLE

This study was funded by the HES-SO University of Applied Sciences and Arts of Western Switzerland. The sponsor had no role in the conception, design, methods, subject recruitment, data collection, data analysis and preparation of this article.

REFERENCES

1. WHO World Health Organization report/ WHO Global report on falls prevention in older age. 2007. ISBN 978 92 4 156353 6, 2007.
2. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis. *BMC geriatrics*. 2014;14:14.
3. Lauretani F, Ticinesi A, Gionti L, Prati B, Nouvenne A, Tana C, et al. Short-Physical Performance Battery (SPPB) score is associated with falls in older outpatients. *Aging clinical and experimental research*. 2018.
4. Hopewell S, Copsey B, Nicolson P, Adedire B, Boniface G, Lamb S. Multifactorial interventions for preventing falls in older people living in the community: a systematic review and meta-analysis of 41 trials and almost 20 000 participants. *Br J Sports Med*. 2019.
5. Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC geriatrics*. 2011;11:33.
6. Liu SW, Obermeyer Z, Chang Y, Shankar KN. Frequency of ED revisits and death among older adults after a fall. *The American journal of emergency medicine*. 2015;33(8):1012-8.
7. Hilliard MJ, Martinez KM, Janssen I, Edwards B, Mille ML, Zhang Y, et al. Lateral balance factors predict future falls in community-living older adults. *Archives of physical medicine and rehabilitation*. 2008;89(9):1708-13.
8. Arvin M, van Dieen JH, Faber GS, Pijnappels M, Hoozemans MJ, Verschueren SM. Hip abductor neuromuscular capacity: A limiting factor in mediolateral balance control in older adults? *Clinical biomechanics (Bristol, Avon)*. 2016;37:27-33.
9. Arvin M, Hoozemans MJ, Burger BJ, Rispens SM, Verschueren SM, van Dieen JH, et al. Effects of hip abductor muscle fatigue on gait control and hip position sense in healthy older adults. *Gait & posture*. 2015;42(4):545-9.
10. Cofre Lizama LE, Pijnappels M, Faber GH, Reeves PN, Verschueren SM, van Dieen JH. Age effects on mediolateral balance control. *PloS one*. 2014;9(10):e110757.
11. Nankaku M, Kanzaki H, Tsuboyama T, Nakamura T. Evaluation of hip fracture risk in relation to fall direction. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2005;16(11):1315-20.
12. Hayes WC, Myers ER, Morris JN, Gerhart TN, Yett HS, Lipsitz LA. Impact near the hip dominates fracture risk in elderly nursing home residents who fall. *Calcified tissue international*. 1993;52(3):192-8.
13. Greenspan SL, Myers ER, Kiel DP, Parker RA, Hayes WC, Resnick NM. Fall direction, bone mineral density, and function: risk factors for hip fracture in frail nursing home elderly. *The American journal of medicine*. 1998;104(6):539-45.
14. Choi WJ, Cripston PA, Robinovitch SN. Effects of hip abductor muscle forces and knee boundary conditions on femoral neck stresses during simulated falls. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2015;26(1):291-301.
15. Gafner SC, Bastiaenen CH, Ferrari S, Gold G, Terrier P, Hilfiker R, et al. Hip muscle and hand-grip strength to differentiate between older fallers and non-fallers: a cross-sectional validity study. *Clinical interventions in aging*. 2018;13:1-8.
16. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ (Clinical research ed)*. 2015;351:h5527.
17. Fletcher RH, Fletcher SW, Fletcher GS. *Clinical epidemiology : the essentials*. Fifth edition: Lippincott Williams & Wilkins, a Wolters Kluwer business; 2015.
18. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E, et al. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. *Journal of geriatric physical therapy (2001)*. 2017;40(1):1-36.
19. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). *Age and ageing*. 2005;34(6):614-9.
20. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975;12(3):189-98.
21. 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. *Journal of gerontology*. 1994;49(2):M85-94.
22. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Physical therapy*. 2000;80(9):896-903.
23. Gafner S, Bastiaenen CHG, Terrier P, Punt I, Ferrari S, Gold G, et al. Evaluation of hip abductor and adductor strength in the elderly: a reliability study. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity*. 2017;14:5.
24. Krause DA, Schlagel SJ, Stember BM, Zoetewey JE, Hollman JH. Influence of lever arm and stabilization on measures of hip abduction and adduction torque obtained by hand-held dynamometry. *Archives of physical medicine and rehabilitation*. 2007;88(1):37-42.
25. Morcelli MH, Rossi DM, Karuka AH, Crozara LF, Hallal CZ, Marques NR, et al. Peak torque, reaction time, and rate of torque development of hip abductors and adductors of older women. *Physiotherapy theory and practice*. 2016;32(1):45-52.

26. Allet L, Kim H, Ashton-Miller J, De Mott T, Richardson JK. Frontal plane hip and ankle sensorimotor function, not age, predicts unipedal stance time. *Muscle & nerve*. 2012;45(4):578-85.
27. Pamoukdjian F, Paillaud E, Zelek L, Laurent M, Levy V, Landre T, et al. Measurement of gait speed in older adults to identify complications associated with frailty: A systematic review. *Journal of geriatric oncology*. 2015;6(6):484-96.
28. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*. 1991;39(2):142-8.
29. Palmer TB, Thiele RM, Williams KB, Adams BM, Akehi K, Smith DB, et al. The identification of fall history using maximal and rapid isometric torque characteristics of the hip extensors in healthy, recreationally active elderly females: a preliminary investigation. *Aging clinical and experimental research*. 2015;27(4):431-8.
30. Bujang MA, Adnan TH. Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis. *Journal of clinical and diagnostic research : JCDR*. 2016;10(10):Ye01-ye6.
31. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York Wiley; 2013.
32. Bento PC, Pereira G, Ugrinowitsch C, Rodacki AL. Peak torque and rate of torque development in elderly with and without fall history. *Clinical biomechanics (Bristol, Avon)*. 2010;25(5):450-4.
33. Ward RE, Leveille SG, Beauchamp MK, Trivison T, Alexander N, Jette AM, et al. Functional performance as a predictor of injurious falls in older adults. *Journal of the American Geriatrics Society*. 2015;63(2):315-20.
34. Hars M, Audet MC, Herrmann F, De Chasse J, Rizzoli R, Reny JL, et al. Functional Performances on Admission Predict In-hospital Falls, Injurious Falls and Fractures in Older Patients: a Prospective Study. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2018.

CHAPTER 6

Reliability and validity of an adapted hip abductor strength measure in a closed chain position. A new fall risk assessment for older persons? A study protocol

Simone C. Gafner, Caroline H. G. Bastiaenen, Emmanuel Biver, Serge Ferrari,
Lara Allet

BMC geriatrics. 2021;21(1):110.
DOI: 10.1186/s12877-021-02048-6

ABSTRACT

Background

Persons aged ≥ 65 years are currently the world's fastest growing age group. An important complication of age is the increasing risk of falls. Falls have multifactorial etiology and modifiable risk factors open for interventions in prevention and rehabilitation, are of high interest. In this context, strong hip abductors seem to be important to prevent falls. A newly adapted measurement device to measure hip abductor strength (HAS) in a closed chain position was developed. We aim to assess feasibility, intra- and inter-tester reliability and construct and criterion validity of the new measure.

Methods

In two subsequent parts a feasibility, reliability and validity study with an adapted measurement instrument for the assessment of HAS (index test) in a closed chain position in persons aged ≥ 65 years will be conducted. Part I investigates feasibility of the measure in clinical settings as well as reliability of the new HAS test ($n=26$). Part II evaluates construct and criterion validity ($n=169$). Construct validity will be assessed cross-sectional, criterion validity by comparison with prospectively followed up fall history for 12 months (external criterion) and other functional fall risk assessments (Short Physical Performance Battery, Timed Up and Go test, usual gait speed and hand grip strength).

Discussion

Results of feasibility, will give insight in its applicability in daily clinical life and clinimetric properties will show if measurements of HAS in a closed chain position should be encouraged to include in fall risk assessments in older adults.

BACKGROUND

Persons aged ≥ 65 years are currently the world's fastest-growing age group (1). The United Nations state that one in six persons in the world and one in four persons living in Europe and Northern America will be aged over 65 by 2050 (1). They also predict that the number of persons over 80 years is going to triple, from 143 million in 2019 to 426 million in 2050 (1).

An important complication of age is the increased risk of falls. About one-third of community dwellings over 65 years and up to half of those in long-term care institutions fall at least once a year (2). Those impressive numbers show that fall risk detection and prevention is a global public health challenge and should have high priority. Falls can cause serious consequences such as loss of independence (3) and increased mortality (4). Additionally to a fall itself, older persons are in a high risk to enter in a vicious cycle of deconditioning after a fall (5). A first fall and related reduced physical conditions combined with fear of falling might lead to insecurity and to even more falls (5). It is of utmost importance older persons at risk of falls could be detected as soon as possible, to support them in preventing falls and postponing recurrent falls with the aim to reinforce them to reduce serious personal consequences.

Falls have a multifactorial etiology and risk factors can be categorized in environmental, psychological, or cognitive factors, as others linked to medication and physiological aspects (6). Of high interest are modifiable risk factors open for interventions in prevention and rehabilitation. Compromised mobility, balance as well as diminished strength of the lower extremities are within this context mentioned in a systematic review of Sousa et al.(6).

Functional lower-extremity weakness has been shown important to enhance the risk of the transition from a non-faller into a faller (7). Hip abductor muscle strength shows a more pronounced decline in older persons than hip extensor muscle strength of the leg (8). The hip abductor muscle group particularly also is perceived to be important for older persons because of their significant role in maintaining mediolateral balance control (9-11) and their relation to lateral and posterolateral falls (12). Furthermore, hip frontal plane muscle strength and spinal moments of force are important parameters to counteract destabilization in the mediolateral direction during single-leg support and stabilizing head, arms and trunk over the support leg and thus prevent the center of mass to rapidly fall downward and lateral toward the unsupported swing side (13). At the actual moment of impact of an acute fall, muscle contraction of the hip abductors also seems to act protectively against stress-induced on the femoral neck and likewise the risk of a femoral fracture (14).

In our previous research about hip abductors strength and fall risk of older persons, we found that hip abductors strength measured in a side-lying position of the participant can discriminate quite well between older fallers and non-fallers (15-17). However, the hip abductor strength assessment procedure used in these studies was too long (mean 10.58 ± 1.56 min) for clinical use (16). A side-lying position also is not functional as falls usually occur during walking and climbing stairs (7). Measuring hip abductor strength (HAS) in an adapted position; in a standing, closed chain position in which the hip abductors must help stabilize the upper body over the single-leg support and prevent the lateral shift of the center of mass towards the swing side (13), seems much more functional and therefore interesting.

We developed an adapted measurement device, based on our experiences with our already existing measure in a side-lying position, to measure HAS in a closed chain position.

The objective of the present study is to assess feasibility, intra-, and inter-tester reliability as well as construct and criterion validity of the adapted measure in a standing, closed chain position. If reliable and valid this measure would offer caregivers easily feasible possibilities to assess HAS in a functional position in a daily clinical setting. The inclusion of hip abductors strength in the multifactorial assessment of the risk of falls might help to improve the detection of persons at risk of falls and to shape a well targeted guidance for those persons. Benefits will concern a substantial percentage (~40%) (2) of the permanently growing total number of the population aged ≥65 years.

METHODS AND ANALYSIS

A feasibility, reliability, and validity study with a new measurement instrument for the assessment of hip abductor strength (index test) in a closed chain position (described below), in persons aged ≥ 65 years will be conducted in two subsequent parts.

Part I: Feasibility of the measure in clinical settings as well as the intra- and inter-tester reliability of the new hip abductor strength test will be assessed.

Part II: Construct and criterion validity of the hip abductor strength values, assessed with the new measurement instrument, will be evaluated.

The presented study protocol is a clinimetric study nested within a prospective cohort study about an integrative fragility fracture prediction model called AFFIRM-CT study, including participants aged ≥65 years. The protocol follows the SPIRIT guidelines, which are an international standard for trial protocols and provide recommendations for their content (18). Participants flow through the study (part I and part II) is shown in Figure 1.

Figure 1: Flow of participants through the study. PI Principal Investigator, MMSE Mini Mental State Examination, SPPB Short Physical Performance Battery, GS Gait Speed, HAS Hip Abductor Strength, TUG Timed Up and Go, HGS Hand Grip Strength

The index test; description of the hip abductors strength test used in part I and part II

A force measurement device will be attached to the participant with a belt on the height of the great trochanter of the non-tested leg. The participant will be invited to

stand sideways close to a wall, with their hip in a neutral position in the sagittal and frontal plane (Figure 2). The patient's shoulder should not touch the wall. A chair for stabilization will be placed in front of the tested person.

The participant is invited to raise the non-tested leg (the leg close to the wall) up and flex the knee approximately 20°, in a way that the foot has no floor contact (Figure 2). The tester stands behind the participant and guides the participant's pelvis in direction of the wall by tactile and verbal information until the dynamometer is slightly touching the wall (correct starting position).

The participant then is asked to push with the standing leg (the one further away from the wall) as hard (assessment of the maximum voluntary isometric contraction (MVIS)) and in the second round as fast as possible (assessment of the rate of force generation (RFG)) in the

direction of the wall to squeeze the dynamometer between the pelvis and the wall. Verbal encouragement will be given by saying to "push", "push", "push", ... against the wall over a five second time span. One to three submaximal test trials are used until the participant understands the test movement. The test is repeated three times per leg (6 trials in total for the assessment of MVIS and 6 trials for the RFG of both legs, with a break of 30 seconds between each trial).

Figure 2: Schematic picture of our measurement instrument and our testing position during the measurement procedure.

PART I: INTRA- AND INTER-TESTER RELIABILITY AND FEASIBILITY

Reliability, in the context of this study, is « the extent to which scores for patients who have not changed (between the subsequent test performances) are the same for repeated measurements under several conditions: e.g. by different persons on the same occasion (inter-tester); or by the same persons on different occasions (intra-tester)» (19).

Study Aims and hypotheses:

The aims of part I are the evaluation of the intra- and inter-tester reliability of the results of the measurements, by health care professionals. According to our previous studies, we hypothesize that the intra- and inter-tester reliability will achieve ICC_{agreement} values of 0.8 and above, corresponding to good to excellent reliability (20). We further aim to list some aspects of feasibility of the new testing procedure. We will list if the testing procedure is well tolerated and appreciated by the participants and can be performed in a feasible time frame for clinical use (< 10min).

Study design

Part I is a reliability study within a cross-sectional design. It is the first study evaluating the reliability of the new testing procedure and an important first step for a later implementation in daily clinical use.

Participants in and exclusion criteria

Participants aged ≥65 years, able to walk household distances with or without an assistive device, hospitalized, already participating in the AFFIRM-CT study, and have signed informed consent for this part will be included. We excluded participants with cognitive, orthopedic, or neurological disorders associated with an increased fall risk (e.g. moderate or severe dementia (MMSE <18) (21), participants with stroke, persons with Parkinson's disease, persons who had a hip replacement within the previous year) (6, 22) and persons who presented contraindications for or would adversely affect the strength tests (see our previous publication (17)) .

Procedure

At the moment of enrollment between January 2021 and January 2023, participants of the cohort of the AFFIRM-CT study will be asked to participate in this nested-in study. The study nurse of the cohort will verbally inform potential study participants about this study and distribute written information. One week after having consented to the study nurse to be approached for the study, the principal investigator (PI) will call the potential study participants to ask if they have questions, if they agree to participate (hand-in the signed informed consent) and give their permission to forward already collected relevant participant information collected within the cohort towards the nested-in study. As soon as the PI obtains the informed consent, the study nurse will consult the participants' information and check the respective in- and exclusion criteria (described above) for the designated participant. When he can be included, the study nurse will arrange the first appointment. For the diagram of participants flow through the study, the study nurse will record the characteristics of all interested participants in a coded form.

Socio-demographic data (age, sex, height, and weight) will be extracted from the participant record of the cohort study. In addition, participants' fall history in the last 6 months (6) and the Mini Mental State Examination questionnaire (MMSE) (21) will be assessed by the study nurse. Finally, the study nurse will ask the participant to perform the short physical performance battery test (SPPB) (23, 24) and assess her/ his usual gait speed on a 4m walkway (25, 26). These two functional fall risk assessments are part of the relevant characteristics of the study population.

The inter-tester reliability will be evaluated between two assessors (health care professionals). The exact testing procedure for the index test, hip abductors strength assessment, is described earlier in the method section. For the evaluation of the inter-tester reliability, a concealed allocation of two randomly chosen assessors (health care professionals) will measure the hip abductors strength of a participant with a 5min break between the assessments of the two assessors (see randomization and conceal allocation section below). Similar to instructions in clinical situations, the assessors will get a short 10min instruction of how to use the dynamometer prior to the measurements. The HAS (MVIS and RFG) of the left or right side (randomly chosen) will be evaluated. The patient will be asked to push the dynamometer first as hard as possible (assessment of MVIS) and in a second-round as fast as possible against the wall (assessment of RFG). Every test (MVIS and RFG) will be repeated three times with a break of 30 seconds between each trial. The two assessors will not discuss the collected strength results and will be blinded to the hip abductors' strength test results of the other assessor as well as to all fall risk related parameters (e.g. fall history, results of other functional fall risk assessments).

Participants will then have a break of two hours before the re-assessment of their HAS by the pre-designated assessor for the evaluation of the intra-tester reliability.

The intra-tester reliability will be evaluated by two measurements with the same participant and by the same assessor. The two assessments will be done two hours apart, an offset considered to be close enough so that no health status change regarding the HAS measurement is expected but far enough that the participants can recover between the two test sessions. The assessor will be blinded for all fall risk related parameters.

Directly after the two sessions, participants will be asked to answer a short "feasibility" questionnaire about their perception during the test sessions and their respective motivation to complete the tasks. For the assessment of the feasibility of the new measurement instrument, the time to complete the hip abductors strength test will be recorded in seconds for the three trials per leg (one side, MVIS, and RFG). The assessor additionally will be asked to complete a short questionnaire with multiple choice answer options and open questions about the feasibility of the test in clinical life (comprehensibility of the test instructions for the participant and practical points for the installation of the measurement device (belt, manageability of the device, starting and stopping of measurements)). Participants' also will be asked to complete a questionnaire with closed and open question options about their perception of the test (duration of the test), in case that any pain occurred during or after the test indicate the intensity on the numerating rating scale (NRS) and indicate on a body chart where the pain occurred. Next, there will be some questions about the comprehensibility of the instructions and the test. In both questionnaires, the one for the assessors as well as the one for the participants, a field for comments will allow giving open-ended feedback on all aspects of the measurement feasibility.

Randomization and allocation concealment

For part I of the study, two concealed randomizations will be needed; once for the order of the assessors for the inter-tester reliability and once to decide the order of the tested leg for each patient. The concealed block randomization will be performed, in a block size of four with a validated electronic randomization system, and will be performed by a secretary, completely independent of the study and does not have access to raters nor patients. The allocation concealment will be granted by sequentially numbered, opaque, sealed envelopes one pile with white envelopes for the order of the assessors and one with grey envelopes for the order of the tested leg, which the secretary will assign just before the beginning of the testing procedure.

Data collection

The data collection will be performed as soon as the participant officially gave their permission and signed the informed consent. Participants' information and test results will be entered in the electronic case report form (e-CRF, see "Data handling and storage for part I and part II of the study" section).

Socio-demographic data and population characteristics

The socio-demographic data (age, sex, height, and weight) will be extracted from the electronic participant record of the AFFIRM-CT cohort study.

Participants' fall history in the last 6 months (6) will be obtained by personal questioning with a beforehand explanation of the definition of a fall. A fall is defined as an event resulting in a person inadvertently coming to rest on the ground, floor, or other lower level (2).

Cognitive impairment is known as an influencing factor of the fall risk in older adults (27). Consequently, the study nurse will assess cognitive impairments with the Mini Mental State Examination (MMSE), a validated assessment tool for older persons (27, 28).

Functional fall risk assessments

To complete the data of the participants' characteristics the study nurse will further assess the SPPB (24) and participants' usual gait speed on 4m (25, 26). The SPPB is a functional physical measure of performance for older persons (29) with possible scores ranging from zero (worst) to 12 points (best). For the assessment of the walking speed, participants will be asked to twice walk a 4m distance at their usual speed with a static start (26). The faster of the two trials will be retained for the analyses (26).

Index test; hip abductors strength test

The data collection for the HAS tests is described earlier in the method section. The strength values will be entered in the e-CRF.

Sample size calculation for the reliability of the index measure

According to previous study results assessing HAS (30) we are expecting ICC_{agreement} values of 0.8 and above, corresponding to good to excellent reliability for the intra-

and inter-tester reliability (20). For the 95% confidence intervals, we are expecting values of ± 0.2 which lead us to a sample size of 23, calculated with the statistical environment R (31) (package: ICC sample size), for part I. Considering a 10% drop out rate, we will include 26 participants. All 26 participants will be assessed for inter-, and intra-tester reliability.

Data analysis

Data analyses will be performed with SPSS version 25 (IBM Corp. Released 2017. Armonk, NY: IBM Corp.). Descriptive statistics for the sample characteristics, time to complete the tests, and number of participants able to perform the test will be performed. For continuous variables, we will present the mean and standard deviation (mean ± SD). Intra-tester and inter-tester reliability will be assessed using the ICC_{agreement} (A, 1) model, which assesses the degree of agreement among measurements assuming a two-way random-effects model (15, 32). In addition, we will compute 95% confidence intervals (CI) through bootstrapping (5000 resamples, bias-corrected, and accelerated percentile method) (15). The standard error of measurement (SEM) will be computed as follows: $SEM = \sqrt{(\sigma_{observer}^2 + \sigma_{residual}^2)}$ where $\sigma_{observer}$ is the variance between the two testers and the residual variance, which is partly due to the unique combination of participants and observers (33, 34). The minimal detectable change (MDC) will be calculated as follows: $MDC = 1.96 * SEM * \sqrt{2}$. The MDC will be normalized by the mean and expressed in percentage. We will further use Bland Altman analyses to investigate absolute reliability and determine between session agreements of the strength measurements (35, 36). We will also present the 95% limits of agreement (LoA 95%) representing two SD above and below the mean differences between the sessions (30).

PART II: CONSTRUCT AND CRITERION VALIDITY

The validation will be performed according to the principles of COSMIN (34) and following the checklist of « Standards for Reporting Diagnostic accuracy studies » (STARD) (37). The construct we are interested in, in part II of this study, is measuring the "fall risk" of persons aged ≥ 65 years. As indicated in the introduction, falls and fall risk are influenced by a vast number of influencing factors with a very multifactorial etiology (a multidimensional construct). Therefore, "fall risk" is a rather abstract construct, which researchers and clinicians try to assess indirectly by a variety of different tests. Interesting factors for prevention or rehabilitation interventions for "fall risk" are performance variables like balance and strength, which are modifiable by non-pharmacological interventions of health care professionals (e.g. physiotherapists,

ergotherapists, etc.) and open to improvement. Therefore, in our field “fall risk” is currently assessed by performance tests like the SPPB (24), the timed up and go (TUG) (38), gait speed (GS) (26), as well as hand-grip strength (HGS) as a parameter, assumed to be related to whole-body strength and fall risk (39, 40). These tests will be used to assess the construct validity as they are all assumed to be measuring one or more aspects of the not directly to measure construct “fall risk”.

For the evaluation of the criterion validity, the fall history will be used as the predefined external criterion also called the reference standard or gold standard in this part of the study. It is assumed to correctly classify the participants into fallers or non-fallers. Participants not reporting a fall in the follow-up phase of the study and/or within the last 6 months will be classified as non-fallers (41). Participants that report one or more falls will be classified as fallers (41). Participants were informed that a fall was defined as an event resulting in a person inadvertently coming to rest on the ground, floor, or other lower level (2). Ideally, fall history is assessed in a prospective follow up period where participants can closely be followed and their number of falls regularly assessed.

Study Aims and Hypotheses

For the evaluation of construct validity of the hip abductors' strength measurement, we will compare the hip abductors strength to other currently used functional fall risk assessments like the SPPB (24), TUG (38), GS (25), and HGS (40). We aim to assess the strength and direction of the correlation between currently used fall risk assessments SPPB, TUG, GS, and HGS with the hip abductors strength, each separately. We hypothesize to find moderate (42) positive correlations (0.40-0.69) between hip abductors strength, SPPB, GS, and HGS and moderate negative correlations between the TUG and hip abductors strength. We are expecting moderate correlations as HAS assesses the same construct but not completely the same parameters as the other fall risk assessments. SPPB and TUG, and in some ways GS, include the assessment of functional parameters like gait and balance, whereas HGS and HAS assessment include only strength. We expect the correlations between hip abductors, SPPB, GS, and HGS to be positive and hypothesize to find the highest correlation (close to the upper boundary of a moderate correlation ≤ 0.69) between SPPB and HAS as both include a specific strength assessment of the lower limb. The correlation between HGS and HAS is expected to be rather close to the lower boundary of a moderate correlation (≥ 0.4) as they don't assess the strength of the same body parts. The correlation between the TUG and HAS is hypothesized to be almost as strong as between the SPPB and HAS but due to the scoring of the TUG in a negative direction.

The criterion validity will be evaluated considering HAS as the index test and fall history of the last 6 months as well as falls during the follow-up time of 12 months as the predefined reference standard. We intend to evaluate if hip abductors' strength assessed with the new procedure is a good diagnostic tool to classify older persons into fallers and non-fallers and/or can improve fall risk detection by combining it with other functional fall risk assessments. Therefore, for the criterion validity we aim to investigate the diagnostic accuracy of HAS (MVIS and RFG), by evaluating its area under the curve [AUC], sensitivity [sens], specificity [spec], positive predictive value [PPV], negative predictive value [NPV], and positive and negative likelihood ratio [LR+, LR-] and the respective, clinically relevant cut-off values. We will compare these values, in the same population, with the indicators of diagnostic accuracy for TUG, SPPB, GS and HGS. We will further calculate net sensitivity and specificity of HAS to get insight in its possible combination with our chosen fall risk assessments (TUG, SPPB, GS and HGS). We hypothesize to find a good diagnostic accuracy of HAS assessed with the new procedure and AUC values of ≥ 0.85 (16, 43) with a 95% confidence interval not going below a moderately accurate level of 0.7 (16, 44). The diagnostic accuracy will be comparable to that of the fall risk assessments TUG, SPPB and GS and has a better accuracy than HGS. Additionally we expect a high accuracy of HAS in recognizing fallers in parallel testing with TUG, SPPB, GS, and HGS (17).

Study design

Part II is a validity study with prospective data collection regarding the external criterion (fall history) and a cross-sectional design for the evaluation of the construct validity.

Participants in- and exclusion criteria

People already participating in the AFFIRM-CT study, and having signed informed consent for this part will be included. For part II, community-dwelling and hospitalized participants will be included according to the same inclusion and exclusion criteria as the participants of part I of this study.

Procedure

From January 2021 on, participant recruitment is following the same procedure as described in part I until the scheduling of the first appointment after participants signed the informed consent.

During the first appointment, the study nurse will assess the following fall risk related variables: fall history of the last 6 months (6), poly-medication, and participants' physical activity level assessed with a physical activity tracker worn on the wrist over 7 days. Then a health care professional blinded for participants' fall history will assess the following functional fall risk assessments: SPPB (24), TUG (38), usual GS (25), and HGS (40). Afterward, the second assessor blinded to the test results of the functional fall risk assessments and the fall history of the participants will measure participants' HAS according to the given description of the index test. This assessor will receive prior to the measurements, a short 10min instruction of the use of the dynamometer, similar, to what would be done in a clinical situation. The HAS of the left and right side will be evaluated either by starting with the randomly chosen right or left side. As described in part I, the MVIS will be measured by asking the participants to push as hard as possible with the dynamometer against the wall (3 times per leg with a break of 30 seconds between each trial). Then the measurements are repeated, this time by pushing the dynamometer as fast as possible against the wall (assessment of the RFG).

Once the hip abductors strength (index test) is evaluated, participants will be followed up for a 12 months period to assess their number of falls (reference test). Participants will be provided with fall diaries in which they are asked to write down their falls and the circumstances/reasons of the falls (intrinsic or extrinsic) and occurring injuries (e.g. fractures) every time such an event will happen. Participants' will thoroughly be explained what the definition of a fall is and how to fill out the fall diary. A fall is defined as an event resulting in a person inadvertently coming to rest on the ground, floor, or other lower level (2). Every 6 months, the study nurse will call participants to collect the fall information and provide coaching for the fall diaries if needed.

Data collection

Data collection will be performed as soon as the participant officially gave their permission and signed the informed consent. Participants' information and test results will be entered into the electronic case report form (e-CRF).

Socio-demographic data and fall risk assessment

As for part I, the reliability part, the socio-demographic data (age, sex, height, and weight) will be obtained by a phone call or from the electronic participant record of the AFFIRM-CT study. A variety of non-functional fall risk related variables for the description of our participants will be assessed. The fall history of the last 6 months

and the MMSE (28) will be evaluated by the study nurse exactly as described in the data collection section of part I of this protocol. The number and names of the medication the participants take will be written down on the participant's documents to get an indication of their poly-medication. Further, participants' physical activity level, assessed with a physical activity tracker worn on the wrist over a time span of 7 days, and the use of walking aids/assistive devices will be registered in the e-CRF.

Functional fall risk assessments

Additionally to the SPPB (24) and usual GS (25, 26) assessed exactly as described in part I, the Timed up and Go test (TUG) (38, 45) will be assessed. The time to complete the test at the usual GS will be recorded in seconds (45). The hand-grip strength will be evaluated with a Jamar dynamometer with participants sitting on a chair, their elbow in 90° flexion with the forearm resting in a neutral position on the arm of a chair and the wrist in 0-30° dorsiflexion (40). The strength will be evaluated in kg (40).

Index and reference test assessment

A health care professional/assessor, blinded for the test results of the functional fall risk assessments and the fall history of the participants, will measure participants' HAS following the testing procedure described for the index test. The strength values will be entered in the e-CRF.

During the follow-up phase of 12 months, participants will be asked to complete fall diaries as soon as they experienced a fall. The study nurse additionally will call participants every 6 months.

Sample size calculation

According to our previous findings of the AUC of hip abductors strength in a standing position, good AUC values (0.85) are expected (16, 43) with a 95% confidence interval not going below 0.7 (16). Using the statistical environment R (31) with a significance level of 5% and a power of 80% we need 31 positive and 104 negative cases for the evaluation of the diagnostic accuracy (criterion validity). Compensating for an expected drop-out rate of 20% over the 12 months follow up period, we will have to include 169 participants. As the sample size calculation for the construct validity is smaller than for the criterion validity we base our study on the sample size for the criterion validity.

Data analysis

Data analyses will be performed with SPSS version 25 (IBM Corp. Released 2017. Armonk, NY: IBM Corp.). Descriptive statistics for the sample characteristics and the functional and non-functional fall related variables will be performed. For continuous variables, the mean and standard deviation (mean $\pm$ SD) will be presented.

The definition of the evaluation of the MVIS and RFG of the HAS is described in our previous article (15). MVIS and RFG will be normalized to body weight (16).

For the construct validity, the correlations of the HAS (MVIS and RFG) with the functional fall risk assessments SPPB, TUG, GS, and HGS will be presented.

For the criterion validity, the AUC and the diagnostic values of HAS, TUG, SPPB, GS, and HGS, compared to the chosen external criterion, history of falls, and 12 months fall prediction (sensitivity, specificity, PPV, NPV, LR+, and LR-) as well as cut-off values will be presented. We will further calculate the net sensitivity and net specificity to present the additional use of HAS in a test battery with other fall risk assessments (SPPB, TUG, gait speed). Additionally we will calculate indices for calibration and discrimination.

Data handling and storage for part I and part II of the study

Study participation is voluntary. Handling in the signed informed consent is a requirement before data collection can start. The Swiss legislation will be followed for the research data management and personal data of the participants. All data of study participants will be encrypted by a code. Participants ID will consist of the three letters HAS (hip abductor strength) and a continuous number starting from 001 up to 202 (ex. HAS_001). The coding sheet will be stored separately from all study documents in a cupboard and on a secured server of our institution in a folder only accessible by the principal investigator.

Throughout the entire trial, strict confidentiality is ensured and only a few authorized persons will have access to the study data. Respecting the Swiss law on human research (Federal Act on Research involving Human Beings (HRA)) and its applicable ordinance ClinO KlinV OClin OSRUm, no data that may identify subjects will be provided. According to the local standards for studies with short duration and minimal risk, no monitoring is planned for the nested-in study. Once data collection and analyses are completed, the coding sheet will be destroyed and no link between participants' names and the codes can be made anymore. The data of the study will be stored for a minimum of 10 years in the institutional archiving system Yareta, the Research Data Repository of Geneva's Higher Education Institutions. Within

the limits of the ethical approval, the encrypted data may be used for related future research questions.

Ethics and Dissemination

This trial will be carried out in accordance with Swiss legislation, the Declaration of Helsinki (46), and Good Clinical Practice guidelines. The protocol was submitted to and approved by the ethics committee of Geneva (Commission Cantonal d'éthique de la recherche sur l'être humain (CCER), BASEC No 2019-01327). The project leader is promptly notified if immediate safety and protective measures have to be taken during the conduct of the research project. If a serious event occurs, the research project will be interrupted, and the Ethics Committee notified. Part I and part II of this study is nested within the AFFIRM-CT study and will be registered in ClinicalTrials.gov. After data collection and analyses are completed, the data of part I and part II will each lead to publications in international peer-reviewed journals.

DISCUSSION

The detection of persons at risk of falls is a major public health concern and should have high priority. Unfortunately, the diagnostic accuracy of currently used functional tests (e.g. SPPB and TUG) to assess older persons' fall risks remains moderate and additionally, those tests usually don't reveal directly targetable parameters for the prevention or rehabilitation of persons at risk. Amongst others, this led to the development of the above-mentioned adapted measurement instrument for hip abductors strength of older persons aged 65 years and over. Understanding the construct of older persons' fall risk might benefit from the additional assessment of HAS as it is evaluating a complementary aspect of falls (17). To perform several tests either in parallel or serial testing increases the certainty of detecting as many persons at risk of falls (true positives or true negatives) as possible (increase in net sensitivity and/or net specificity). As we know that the diagnostic accuracy of currently used functional fall risk assessments is only moderate, we are interested to investigate if HAS could improve a reliable and valid fall risk detection by its use in a test battery. Compared to the currently used functional performance tests to detect the fall risk (e.g. SPPB, TUG) HAS might be a directly targetable fall-risk parameter, which could contribute to treatment goal setting and might also to a well-targeted treatment (to be proven in future studies).

As there are not yet any articles available using the parameter hip abductors strength in a closed chain position, we need to investigate its clinimetric properties before a wide implementation can be considered. We will investigate this in a

population of older adults with no additional diseases known to increase the fall risk. However, many diseases are known to increase the fall risk (e.g. neurological diseases) and future studies might investigate the additional use of the assessment of hip abductors strength for the assessment and treatment of these persons when reliability and validity are proven and acceptable in this study.

Regarding the clinical use of the measurement device for the hip abductors, we are expecting that the measurements are easily feasible in a practical time frame for clinical use, that its reliability for intra- and inter-tester use is high and it shows a good validity for the construct “fall risk” compared to the other selected measures and the differentiation of fallers and non-fallers compared to fall history).

Following the STARD guidelines (37), the methodological aspects important for a diagnostic study were specifically taken care of in this study. First, the selection of the population will be performed consecutively and corresponds to the specific settings in which the HAS measure will be used. The in-and exclusion criteria lead to a representative sample of the target population. The data collection is defined before the index and reference test will be performed and the assessor of the index test will be blinded for the results of the reference test. The nested nature of this study will not bias the study results of either study (reliability and validity part as well as the AFFIRM-CT study) as for both studies no intervention is planned as well as the included population is in line with the population in which the measurement instrument is meant to be used. The aim of the AFFIRM-CT study is to develop and validate an integrative fragility fracture prediction model. The model will integrate fall risk, bone strength, and fall biomechanics to predict a subject's risk of fracture. Successful development and proper validation of this integrative fragility fracture prediction model will provide a better estimation of an individual's fall and fracture risk and will allow broader screening across the growing population aged ≥ 65 years. The advantage of our study nested within the AFFIRM-CT study is that we get access to a high number of participants and that we investigate in a subsequent step if a combination of functional fall risk assessments (SPPB, TUG, GS, HGS, and HAS) with more medical information (e.g. bone density, impact load) can be advantageous to identify older persons' at risk of falling. The information about the validity of the HAS measurement, on the other hand, will add to the precision of the prediction model of the AFFIRM-CT study. The functional testing position and clinically easily applicable test of the HAS measurement, as well as the 12 months prospective assessment of the numbers of falls of our participants, is a further big advantage of our study.

The blinding of the persons assessing hip abductors strength, for participants' fall history and functional fall risk assessment test results, will be performed with great care. It is important that the assessor cannot intentionally influence the verbal encouragement during the strength testing of fallers or non-fallers according to their

fall history for example. Additionally, we randomize the order of the assessors in the reliability part to simulate a situation as close to practical situations as possible. However, the randomization of the testing order for the functional fall risk assessments as well as the assessment of the MVIS before the RFG is not possible due to practical reasons at the testing location and to the nested nature of the presented study. Therefore, we will statistically evaluate if learning effects occurred and carefully check that the resting time of 5 min between the functional tests will be respected. Further, the statistician will be blinded for the outcomes of fallers and non-fallers for the analyses. To our knowledge, our measurement device is the first instrument easily accessible and transportable for clinical practice that is measuring HAS in a closed chain position. During the development process, it was of great importance that the developed measurement device is easily applicable in the daily clinical life of health care professionals. With the collected information about feasibility, we will get insight if this goal was achieved and the clinimetric properties will show if the measurement of HAS in a closed chain position should be encouraged as a reliable and valid measure for fall risk assessment in older adults.

ABBREVIATIONS

ABD: hip abductor; AUC: area under the curve; CCER: Commission Cantonal d'éthique de la recherche sur l'être humain (ethical commission); CI: confidence interval; e-CRF: electronic case report form; GS: gait speed; HAS: hip abductor strength; HGS: hand grip strength; LoA: limits of agreement; LR+: positive likelihood ratio; LR-: negative likelihood ratio; MMSE: Mini Mental State Examination; MVIS: maximum voluntary isometric strength; NL: Netherlands; NPV: negative predictive value; NRS: numerating rating scale; PI: principal investigator; PPV: positive predictive value; RFG: rate of force generation; SNF: Swiss National Science Foundation; SPPB: Short Physical Performance Battery; STARD: standard of reporting diagnostic accuracy studies; TUG: Timed Up and Go

DECLARATIONS

Ethics approval and consent to participate:

The protocol was submitted to and approved by the ethics committee of Geneva (Commission Cantonal d'éthique de la recherche sur l'être humain (CCER), BASEC Number 2019-01327). This trial will be carried out in accordance with Swiss legislation, the Declaration of Helsinki and Good Clinical Practice guidelines. Study participation

is voluntary. Handling in the signed informed consent is a requirement before data collection can start.

CONSENT FOR PUBLICATION

Not applicable

AVAILABILITY OF DATA

Not applicable

COMPETING INTEREST

The authors declare that they have no competing interests.

FUNDING

This study was peer reviewed and funded by the Swiss National Science Foundation (SNF). We declare that the SNF was not involved in designing the study, data collection, analyses, and interpretation of the data and writing of the manuscript.

AUTHORS' CONTRIBUTION

All authors (SCG, CHGB, EB, SF, and LA) were fully involved in the conception and design of the study, in drafting the manuscript and critically revising it. All the authors read and approved the final submitted version. **SCG:** study conception and design, initial draft and final approval of the submitted version of the article; **CHGB:** study conception and design, writing the draft and final approval of the submitted version of the article; **EB:** study conception and design, and final approval of the submitted version of the article; **SF :** study conception and design, and final approval of the submitted version of the article; **LA:** study conception and design, writing draft and final approval of the submitted version of the article.

ACKNOWLEDGMENT

The authors would like to thank Eugene Rameckers, assistant professor of the Department of Rehabilitation Medicine at Maastricht University (NL), for his valuable inputs in the first stages of the theoretical development of the adapted measurement instrument.

REFERENCES

1. United Nations DoEaSA, Population Division (2019). United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423).
2. WHO World Health Organization report/ WHO Global report on falls prevention in older age. 2007. ISBN 978 92 4 156353 6, 2007.
3. Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. BMC geriatrics. 2011;11:33.
4. Liu SW, Obermeyer Z, Chang Y, Shankar KN. Frequency of ED revisits and death among older adults after a fall. The American journal of emergency medicine. 2015;33(8):1012-8.
5. Cumming RG, Salkeld G, Thomas M, Szonyi G. Prospective study of the impact of fear of falling on activities of daily living, SF-36 scores, and nursing home admission. The journals of gerontology Series A, Biological sciences and medical sciences. 2000;55(5):M299-305.
6. Sousa LM, Marques-Vieira CM, Caldevilla MN, Henriques CM, Severino SS, Caldeira SM. Risk for falls among community-dwelling older people: systematic literature review. Revista gaucha de enfermagem. 2017;37(4):e55030.
7. Muir SW, Berg K, Chesworth BM, Klar N, Speechley M. Modifiable Risk Factors Identify People Who Transition from Non-fallers to Fallers in Community-Dwelling Older Adults: A Prospective Study. Physiotherapy Canada Physiotherapie Canada. 2010;62(4):358-67.
8. Daun F, Kibele A. Different strength declines in leg primary movers versus stabilizers across age- Implications for the risk of falls in older adults? PloS one. 2019;14(3):e0213361.
9. Hilliard MJ, Martinez KM, Janssen I, Edwards B, Mille ML, Zhang Y, et al. Lateral balance factors predict future falls in community-living older adults. Archives of physical medicine and rehabilitation. 2008;89(9):1708-13.
10. Arvin M, van Dieen JH, Faber GS, Pijnappels M, Hoozemans MJ, Verschueren SM. Hip abductor neuromuscular capacity: A limiting factor in mediolateral balance control in older adults? Clinical biomechanics (Bristol, Avon). 2016;37:27-33.
11. Arvin M, Hoozemans MJ, Burger BJ, Rispens SM, Verschueren SM, van Dieen JH, et al. Effects of hip abductor muscle fatigue on gait control and hip position sense in healthy older adults. Gait & posture. 2015;42(4):545-9.
12. Cofre Lizama LE, Pijnappels M, Faber GH, Reeves PN, Verschueren SM, van Dieen JH. Age effects on mediolateral balance control. PloS one. 2014;9(10):e110757.
13. Johnson ME, Mille ML, Martinez KM, Crombie G, Rogers MW. Age-related changes in hip abductor and adductor joint torques. Archives of physical medicine and rehabilitation. 2004;85(4):593-7.

14. Choi WJ, Crompton PA, Robinovitch SN. Effects of hip abductor muscle forces and knee boundary conditions on femoral neck stresses during simulated falls. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2015;26(1):291-301.
15. Gafner S, Bastiaenen CHG, Terrier P, Punt I, Ferrari S, Gold G, et al. Evaluation of hip abductor and adductor strength in the elderly: a reliability study. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity*. 2017;14:5.
16. Gafner SC, Bastiaenen CH, Ferrari S, Gold G, Terrier P, Hilfiker R, et al. Hip muscle and hand-grip strength to differentiate between older fallers and non-fallers: a cross-sectional validity study. *Clinical interventions in aging*. 2018;13:1-8.
17. Gafner SC, Bastiaenen CHG, Ferrari S, Gold G, Trombetti A, Terrier P, et al. The Role of Hip Abductor Strength in Identifying Older Persons at Risk of Falls: A Diagnostic Accuracy Study. *Clinical interventions in aging*. 2020;15:645-54.
18. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ : British Medical Journal*. 2013;346:e7586.
19. 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. *Journal of clinical epidemiology*. 2010;63(7):737-45.
20. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-63.
21. Monroe T, Carter M. Using the Folstein Mini Mental State Exam (MMSE) to explore methodological issues in cognitive aging research. *Eur J Ageing*. 2012;9(3):265-74.
22. Lo CWT, Tsang WWN, Yan CH, Lord SR, Hill KD, Wong AYL. Risk factors for falls in patients with total hip arthroplasty and total knee arthroplasty: a systematic review and meta-analysis. *Osteoarthritis Cartilage*. 2019;27(7):979-93.
23. Veronese N, Bolzetta F, Toffanello ED, Zambon S, De Rui M, Perissinotto E, et al. Association between Short Physical Performance Battery and falls in older people: the Progetto Veneto Anziani Study. *Rejuvenation research*. 2014;17(3):276-84.
24. 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. *Journal of gerontology*. 1994;49(2):M85-94.
25. Kyrdalen IL, Thingstad P, Sandvik L, Ormstad H. Associations between gait speed and well-known fall risk factors among community-dwelling older adults. *Physiotherapy research international : the journal for researchers and clinicians in physical therapy*. 2019;24(1):e1743.
26. Quach L, Galica AM, Jones RN, Procter-Gray E, Manor B, Hannan MT, et al. The Nonlinear Relationship Between Gait Speed and Falls: The Maintenance of Balance, Independent Living, Intellect, and Zest in the Elderly of Boston Study. *Journal of the American Geriatrics Society*. 2011;59(6):1069-73.
27. Muir SW, Gopaul K, Montero Odasso MM. The role of cognitive impairment in fall risk among older adults: a systematic review and meta-analysis. *Age and ageing*. 2012;41(3):299-308.
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975;12(3):189-98.
29. Freiburger E, de Vreede P, Schoene D, Rydwick E, Mueller V, Frandin K, et al. Performance-based physical function in older community-dwelling persons: a systematic review of instruments. *Age and ageing*. 2012;41(6):712-21.
30. Bruyneel AV, Gafner SC, Ferrari S, Gold G, Monnin D, Terrier P, et al. Intra-rater reliability of hip abductor isometric strength testing in a standing position in older fallers and non-fallers. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity*. 2018;15:9.
31. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
32. McGraw KO, Wong SP. "Forming inferences about some intraclass correlations coefficients": Correction. *Psychological Methods*. 1996;1(4):390-.
33. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *Journal of clinical epidemiology*. 2006;59(10):1033-9.
34. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology)*: Cambridge: Cambridge University Press.; 2011.
35. Seko T, Kumamoto T, Miura S, Kobayashi T, Takahashi Y, Kaneko R, et al. Measuring seated hip extensor strength using a handheld dynamometer: an examination of the reliability and validity of the protocol. *Journal of physical therapy science*. 2015;27(7):2179-82.
36. Awwad DH, Buckley JD, Thomson RL, O'Connor M, Carbone TA, Chehade MJ. Testing the Hip Abductor Muscle Strength of Older Persons Using a Handheld Dynamometer. *Geriatr Orthop Surg Rehabil*. 2017;8(3):166-72.
37. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ (Clinical research ed)*. 2015;351:h5527.
38. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Physical therapy*. 2000;80(9):896-903.
39. Pijnappels M, van der Burg PJ, Reeves ND, van Dieen JH. Identification of elderly fallers by muscle strength measures. *European journal of applied physiology*. 2008;102(5):585-92.

40. Roberts HC, Denison HJ, Martin HJ, Patel HP, Syddall H, Cooper C, et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. *Age and ageing*. 2011;40(4):423-9.
41. Palmer TB, Thiele RM, Williams KB, Adams BM, Akehi K, Smith DB, et al. The identification of fall history using maximal and rapid isometric torque characteristics of the hip extensors in healthy, recreationally active elderly females: a preliminary investigation. *Aging clinical and experimental research*. 2015;27(4):431-8.
42. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg*. 2018;126(5):1763-8.
43. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *Journal of pediatric psychology*. 2014;39(2):204-21.
44. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine*. 2000;45(1-2):23-41.
45. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*. 1991;39(2):142-8.
46. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*. 2013;310(20):2191-4.

CHAPTER 7

Reliability and diagnostic accuracy of commonly used
performance tests relative to fall history in older persons:
a systematic review

Simone C. Gafner, Lara Allet, Roger Hilfiker, Caroline H. G. Bastiaenen

Clinical interventions in aging, 2021;16:1591-1616.

DOI: 10.2147/CIA.S322506

ABSTRACT

Background/Objectives

Early detection of fall risk is crucial for targeted fall prevention and rehabilitation. This systematic review facilitates decision making concerning the optimal choice for a suitable fall risk assessment test for older persons in four different settings. This systematic review provides an overview of reliability and diagnostic accuracy (sensitivity [SE], specificity, likelihood ratios, and post-test probabilities) of commonly used performance measures to assess older persons' fall risk.

Methods

Two reviewers independently screened titles, abstracts, and full texts of all articles identified through a systematic search on the PubMed, Embase, CINAHL, and Web of Science databases. The methodological quality was critically appraised using the Quality Assessment of Diagnostic Accuracy Studies-2 tool and the COSMIN risk of bias tool. Due to a setting-dependent fall risk, four relevant settings were selected for the analyses: 1) community dwelling, 2) nursing home, 3) hospital, and 4) the combined setting of hospitalized and community-dwelling persons.

Results

Recommendations for diagnostic accuracy can be made for the specific settings: Setting 1) Mini-Balance Evaluation System Test, Timed Up and Go test with fast speed (TUG FAST), cognitive TUG FAST, Functional Reach Test, and fast Gait Speed; 2) TUG at usual speed; 3) none of the included tests; and 4) some hip muscle strength and hand grip strength tests might be recommended.

Conclusion

The review's results are applicable for use in clinical practice, both on a population and on an individual level for patients and caregivers. Most measures' reliability is sufficient. However, the included tests' discriminative ability is not optimal for clinical use in relevant settings. The low sample sizes of most of the included studies and a limited amount of studies in all but the community-dwelling setting hinder us from making strong recommendations.

Keywords: Aged, falls, criterion validity, post-test probability, clinical application

INTRODUCTION

Demographic changes are leading to a drastically increasing number of persons aged 65 years and older, which is the world's fastest-growing age group (1). The United Nations states in its World Populations Prospects (1) that one in six persons in the world and one in four persons living in Europe and Northern America will be over age 65 by 2050. They also project that the number of persons over age 80 is going to triple from 143 million in 2019 to 426 million in 2050 (1). This implicates that health care professionals will have an increasing number of older patients in their care and practices in the near future. Considering 28% to 35% of community dwellings for persons over 65 years and 30–50% of those in long-term care institutions fall at least once a year (2), it is clear that taking action is a matter of priority. This high fall rate is particularly disconcerting due to the falls' serious consequences such as loss of independence (3), fear of falling (4), depression (4), and increased mortality (5). Health care professionals should thus aim to detect persons at risk of falls as soon as possible to be able to prevent, if possible, a first fall, or thereafter prevent or at least postpone recurrent falls. To do so, easy-to-use clinical fall risk assessments with high diagnostic accuracy and good reliability are essential.

Falls occur for multifactorial reasons and the number of fall risk assessments is as numerous as these influencing factors (6). Physical performance measures, such as lower limb strength assessments, balance performance tests, and assessing gait speed play an important role in fall risk detection, as older persons often fall while moving around (eg while standing up, walking, or turning around) (7-9). Some examples of commonly used single-performance tests are gait speed (10), one leg stance (OLS) (time that an older person can stand on one leg without arm support) (11, 12), and the functional reach test (FRT) (the maximum reach distance beyond the arm's length while maintaining a fixed support base in the standing position) (13). Examples of commonly used multidimensional performance tests to assess fall risk are the Performance Oriented Mobility Assessment (POMA) (14), the Timed Up and Go test (TUG) (8), and the Short Physical Performance Battery (SPPB) (15).

The quality of a measurement instrument, as the "Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN)" group defines it, depends on the reliability and validity of a measurement instrument, among other factors (16). Reliability, as a test's important clinimetric parameter, and a prerequisite for a good criterion validity to evaluate the diagnostic accuracy, expresses the degree to which a measurement tool is free from measurement error (17). According to the COSMIN group, criterion validity is the degree to which a measurement instrument's scores are an adequate reflection of a gold standard, also called a reference standard (16, 18). The diagnostic accuracy of a measurement

instrument (eg sensitivity [SE], specificity [SP], positive likelihood ratio [LR+], negative likelihood ratio [LR-], and positive and negative post-test probability) is part of the criterion validity. Therefore, the decision of which instrument to choose for a specific fall risk assessment should be influenced by a) the reliability as a prerequisite for validity, b) its intrinsic ability to detect persons at risk or not at risk of falls (sensitivity and specificity), and c) the increase of certainty about this risk after having performed the test (positive and negative post-test probability). Post-test probability is calculated by using the positive or negative likelihood ratios (LR+, LR-) (6) and is very dependent on the fall risk prevalence in the tested population. LRs are calculated as a ratio of SE and SP, and are not dependent on prevalence, except for spectrum bias (19).

Several systematic reviews have been published that assess the reliability or validity of specific fall risk assessments or combinations of multiple fall risk assessment tests (8, 20-26). However, only one systematic review provided post-test probabilities for different fall risk assessments in community-dwelling participants (6). Since then, clinics have been introduced to new tests and many new articles about fall risk tests have been published during the last ten years. In a clinical setting, health care professionals are confronted not only with community-dwelling persons, but also patients living in nursing homes or hospitalized patients.

This systematic review provides an overview of commonly used performance measures to assess the fall risk of persons over 65 years. We present the measures' most clinical relevant clinimetric abilities, a) reliability and, b) criterion validity, which allows the expression of the diagnostic accuracy, including sensitivity and specificity, the likelihood ratios, and the post-test probabilities. This review's results will improve the decision-making process when choosing the most suitable fall risk assessment measure or measurement battery for an older (sub-) population aged at 65 years and older as well as support the individual decision-making process between patient and caregiver.

METHODS

This study was conducted in line with the Joanna Briggs Institute (JBI) Reviewer's Manual for diagnostic accuracy studies (27), the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines (28), and the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement (29). The protocol is registered in the International Prospective Register of Systematic Reviews (PROSPERO), registration number CRD42020157160.

Before presenting the methodology of how the systematic review was performed and its results, the taxonomy, definitions, and methodological aspects of clinimetric studies are explained in the following paragraphs.

As described in the introduction, reliability and validity are important and well-defined clinimetric parameters of a test (16, 17, 30). For the clinical use of measurement instruments, health care professionals should consider carefully the following points while choosing an appropriate test.

- I. The reliability, as a prerequisite for validity, is important. When the reliability is sufficient (eg as judged by the criteria of the COSMIN group (17)) a further investigation of the validity has to follow.
- II. Regarding a first part of the validity and before a measure's results are available, the SE and SP of a measure (intrinsic qualities of a test to detect persons at risk) have to be considered. SE and SP show a measurement instrument's ability to detect true positives and true negatives, respectively (6). As they are calculated only within the diseased (SE) or non-diseased (SP) part of a study population, these values are not as dependent on a disease's prevalence (19). It is interesting to know that, for example, high sensitivity indicates the test's ability to identify people with the diagnosis, but this also means that a negative result for a test with high sensitivity helps to rule out the disease (6).
- III. Once a test is performed and results are available, clinicians must determine if the patient is at risk of falls given the results of the used test. At this point, a test's intrinsic quality is less important because the focus now has to be on the likelihood ratios (LRs) and the post-test probabilities. For example, community-dwelling older persons have a yearly fall risk between 28% and 35% (2). Therefore, to have a valuable test, its results must indicate with a higher certainty than 28% to 35% if a person is at risk of falls (positive post-test probability) or with a lower risk (negative post-test probability). The tested person must know the chance to which a clinician is giving them the right answer (eg at risk of falls or not at risk). The positive post-test probability indicates how high the probability of the disease (fall risk in this systematic review) is after a positive test result is obtained. The negative post-test probability indicates the probability a person has of falling even with a negative test. Post-test probability can be calculated by using positive or negative likelihood ratios (LR+, LR-) (6) and is dependent on the fall risk in the tested population. Unfortunately, likelihood ratios are not always available, and not all health care professionals know how to interpret them in daily clinical use.
- IV. LRs are calculated based on SE and SP, therefore they are presented as a ratio independent from prevalence (except for spectrum bias)(19). A LR+ indicates the increased likelihood of a positive test result for people with the disease than for people without the disease (6, 19). LR- indicates how much less likely a negative test result is for people with the disease than for people without the disease (6, 19). The larger the value is above one for LR+, or the smaller the value is below one for LR-, then the test result is more valuable (19). Using the

Fagan nomogram (Figure 1), likelihood ratios (i.e. the LR+ for a positive test and the LR- for a negative test) allow calculating the post-test probability for every pre-test probability (i.e. every prevalence of the disease)(19, 31). Thus, clinicians need to know the disease's prevalence in their patients, especially in their specific setting, and then they can draw a straight line between the prevalence and the respective LR, thus evaluating the test's post-test probability (19). Given the wide spectrum of patients, settings, and their associated fall prevalence, all of which confront health care professionals in their daily clinical practice, a clear overview of likelihood ratios is required. Due to the high impact of the fall risk prevalence on the post-test probability, a separate presentation of the diagnostic accuracy for the specific settings is important. The following fall risk prevalence can be found in the literature for the different settings: approximately 30% for community dwellings (2), between 30–50% for nursing homes (2), and approximately 24% for hospitalized persons (32, 33). Figure 2 shows an example of the relationship between the prevalence (pre-test probability) and the post-test probability for different cut-off values and fall risk prevalence of the TUG FAST. The dotted line indicates the 30% fall risk prevalence for the community-dwelling setting and the related post-test probability (y-axis) of a positive test, the continuous line indicates the same for a negative test (Figure 2). In the case that only a test's SE and SP are presented in an article, different online diagnostic test calculators exist (eg <http://araw.mede.uic.edu/cgi-bin/testcalc.pl>) and allow clinicians to calculate the LRs and post-test probability for the same setting as for which they chose their article.

Our recommendations of fall risk tests are based on the above-mentioned values and will help facilitate clinicians' choice for an appropriate test in their specific setting.

Figure 1 : Fagan Nomogram to calculate the post-test probability for the TUG FAST in the community-dwelling setting, by drawing a line between the fall risk (pre-test probability, 30% in community-dwelling persons) and the likelihood ratio for a positive (green) or a negative test (red)

Figure 2: Relationship between the prevalence (pre-test probability) and the post-test probability for different cut-off values and fall risk prevalence of the TUG FAST. The dotted line indicates the 30% fall risk prevalence for the community-dwelling setting and the related post-test probability (y-axis) of a positive test and the continuous line for a negative test respectively

Search strategy:

We searched the following databases from their inception until January 31, 2019: PubMed, Embase, CINAHL, and Web of Science. The search terms were grouped by: a) population (aged ≥ 65 years), b) construct (fall risk assessed with single or multidimensional performance tests), c) sensitive search filter according to Terwee et al. (34) to identify studies about the measurement instruments' measurement properties, and d) the exclusion filter that Terwee et al. (34) also proposed. The search terms were combined with Boolean AND, and NOT for the exclusion filter (PubMed search strategy, supplementary material 3). A librarian of the Maastricht University approved the search strategy. Articles published in peer-reviewed journals in English, Dutch, French, or German were included.

Study selection and data extraction:

Two independent reviewers (SCG and CHB) screened the titles, abstracts, and full texts that were then included or excluded according to the predefined criteria. Discussions with a third reviewer (LA) solved any disagreements. The search's results are shown in the PRISMA flow diagram (Figure 3). Title, abstract, and full text screening was performed with Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia (www.covidence.org). The articles had to fulfill the following criteria to be included: a) participants' mean age was 65 years or over; b) contained at least one of the pre-defined single-performance or multidimensional fall risk assessments (index test, see below); c) contained information on reliability (inter-rater, intra-rater) of the single-performance or multidimensional fall risk assessments; d) presented information on validity, eg true positives, true negatives, SE, SP, LR+, and LR-, allowing the calculation of 2x2 tables and post-test probabilities; e) specified fall status (none or $\geq$ one) as an outcome variable (prospective) or classification variable (retrospective) as our reference test; and f) presented information on fall history for at least a 6-month period, either following study enrollment (prospective studies) or recall of falls before the study enrollment. We excluded articles including participants with cognitive dysfunctions or with orthopedic or neurological diagnoses associated with elevated fall risk and articles with little evidence of how falls were defined or documented. Further, we used the validated search filter of the COSMIN group (34) and excluded publication types that were unlikely to include information about diagnostic accuracy accordingly.

Reference Test: A history of falls ($\geq$ one) is a good indicator of the predictive fall risk of older persons (35-37) and is an often-used reference standard. Within this study, a fall was defined as an event resulting in a person inadvertently coming to rest on the ground, floor, or other lower level (2).

Index measures: We performed an extensive literature search in the field of fall risk assessment tests in older persons to find experts within Western Europe who have recently published systematic reviews and/or performed important research in the field of fall risk assessment tests and their clinical use. Subsequently, different performance-based fall risk assessment tests (single- and multidimensional performance tests) and their adapted versions were included in this review.

Single-performance tests included: gait speed, one leg stance (OLS), functional reach test (FRT), grip strength (Grip), and lower limb muscle strength.

Multidimensional performance tests included: Berg Balance Scale (BBS); Mini-, Brief- or Complete Balance Evaluation System Test (MiniBEST, BriefBEST or BEST); Short Physical Performance Battery (SPPB); Timed Up and Go test (TUG); and Performance Oriented Mobility Assessment (POMA).

Some of the included tests were performed at different walking speeds, which introduced different cut-off values. They were thus treated and presented in the analyses as separate tests. The addition of FAST indicates that the test was performed at a speed “as fast as possible.” Tests without the addition of FAST were conducted at “usual speed” (eg TUG and TUG FAST).

Figure 3: PRISMA flow diagram of the literature search of this systematic review

Data extraction:

Two reviewers (SCG and CHB) independently extracted data in the predefined and preliminary tested Excel datasheet that was developed following the STARD statement (28) as well as the COSMIN risk of bias tool (17). A third reviewer (LA) cross-checked the results. The extracted data of the index and reference test were created in “2x2 tables” to calculate the likelihood ratios and post-test probabilities and thus evaluate the diagnostic accuracy.

METHODOLOGICAL QUALITY AND OVERALL RATING OF GOOD MEASUREMENT PROPERTIES

Two independent reviewers (SCG and CHB) critically appraised the selected studies for methodological quality using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (38) for the validity studies and the COSMIN risk of bias tool (RoB) for assessing the reliability studies’ methodological quality (17). For the QUADAS-2 tool, each of the four domains of the risk of bias (patient selection, index test, reference standard, and flow and timing) and the first three domains for concerns regarding applicability were judged as “low,” “high,” or “unclear.” If any signaling question is answered with “no” or “unclear” this flags the potential for bias. For the COSMIN boxes of the RoB, each item was rated individually on the 4-point rating scale (17). Subsequently, we obtained an overall assessment score of a given measurement property by taking the lowest score for any of the items in the box (‘worst score counts’ method) (17). A consensus on one final rating for the COSMIN RoB (17) as well as the QUADAS-2 (38) (Figure 4), was reached by discussion (see Supplementary Table 1 for reliability and Supplementary Table 2 for validity). All studies that fulfilled the inclusion and exclusion criteria, regardless of their methodological quality, underwent data extraction and synthesis.

The measurement instrument’s overall rating was evaluated against the criteria for good measurement properties the COSMIN group (17) proposed and was presented in the respective tables of the reliability and areas under the curve (AUC).

Figure 4: Results of the assessment of the methodological quality with the Quality Assessment for Diagnostic Accuracy Studies 2 (QUADAS-2). On the left proportion of studies with low, high or unclear risk of bias. On the right the proportion of the studies with low, high, or unclear concerns regarding applicability

Statistical analyses and data synthesis:

Due to the fall risk documented for different settings and their influence on diagnostic accuracy (approximately 30% for community dwellings (2), between 30–50% for nursing homes, (2) and approximately 24% for hospitalized persons (32, 33)), analyses were conducted and presented within four specific settings. Whenever several studies supported measurement instruments, data were combined to create larger samples to be more representative of the population of persons aged 65 years and older. When different cut-off values were used in the studies supporting the same tests, we calculated summary receiver operating characteristics curves (SROC) (39) to get an indication of the global performance of the included fall risk assessment tests (single or multidimensional performance tests) (40, 41). AUC for the included tests were pooled when two or more articles of the same test were available. These are presented in Table 1.

Table 1: Validity; area under the curve (AUC) for the different settings, summary of findings

Setting	Test	No of studies	AUC (95% CI)	Overall rating (+/ ?/-)
Community dwelling	BBS	5	0.68 (0.62 to 0.73)	-
	BEST	2	0.75 (0.69 to 0.81)	+
	MiniBEST	2	0.79 (0.64 to 0.88)	+
	BriefBEST	1	0.76 (0.64 to 0.85)	+
	POMA BALANCE	1	(-)	?
	POMA GAIT	1	(-)	?
	TUG	8	0.63 (0.59 to 0.66)	-
	TUG FAST	1	0.58 (0.47 to 0.70)	-
	TUG-8FT FAST	1	(-)	?
	TUG MAN FAST	1	(-)	?
	TUG COG FAST	1	0.65 (0.55 to 0.76)	-
	FRT	1	(-)	?
	GAIT SPEED	1	0.69 (0.62 to 0.76)	-
	GAIT SPEED FAST	1	0.71 (0.64 to 0.77)	+
	GRIP	1	(-)	?
	OLS	1	(-)	?
Nursing home	BBS	1	0.76 (0.66 to 0.84)	+
	BEST	1	0.75 (0.57 to 0.87)	+
	MINI BEST	1	0.71 (0.53 to 0.84)	+
	BRIEF BEST	1	0.75 (0.57 to 0.87)	+
	POMA	1	(-)	?
	POMA BALANCE	1	(-)	?
	POMA GAIT	1	(-)	?
	TUG	1	0.57 (0.40 to 0.73)	-

Table 1: Continued.

Setting	Test	No of studies	AUC (95% CI)	Overall rating (+/ ?/-)
Hospital	SIMPLIFIED POMA	1	0.58 (0.52 to 0.64)	-
	TUG	2	0.55 (0.50 to 0.60)	-
	SPPB	1	0.57 (0.52 to 0.62)	-
Hospital and Community-dwelling	GRIP	1	0.65 (0.48 to 0.79)	-
	HIP STRENGTH ABDUCTORS	1	0.83 (0.71 to 0.94)	+
	HIP STRENGTH ABDUCTORS RFG	1	0.79 (0.67 to 0.91)	+
	HIP STRENGTH ADDUCTORS	1	0.77 (0.65-0.90)**	+
	HIP STRENGTH ADDUCTORS RFG		0.71 (0.57-0.84) **	+
	HIP STRENGTH EXTENSORS	1	0.56 (0.40 to 0.71)	-
	HIP STRENGTH EXTENSORS RFG	1	0.61 (0.46 to 0.76)	-
	HIP STRENGTH EXTERNAL ROTATOR	1	0.74 (0.61 to 0.87)	+
	HIP STRENGTH EXTERNAL ROTATORS RFG	1	0.65 (0.51 to 0.80)	-
	HIP STRENGTH FLEXORS	1	0.76 (0.63 to 0.88)	+
	HIP STRENGTH FLEXORS RFG	1	0.68 (0.54 to 0.82)	-
	HIP STRENGTH INTERNAL ROTATORS	1	0.58 (0.42 to 0.73)	-
	HIP STRENGTH INTERNAL ROTATORS RFG	1	0.62 (0.47 to 0.77)	-

AUC: area under the curve; BBS : Berg Balance Scale, BEST : Balance Evaluation System Test, BriefBEST : Brief Balance Evaluation System Test, MiniBEST: Mini Balance Evaluation System Test, CI : 95% confidence interval,

(-): confidence interval not reported, FRT : Functional Reach Test, GAIT SPEED : Gait speed at usual speed, GAIT SPEED FAST : Gait speed as fast as possible without running, GRIP : grip strength, OLS : One Leg Stance, POMA: Performance Oriented Mobility Assessment, POMA Balance : Performance Oriented Mobility Assessment Balance part, POMA GAIT : Performance Oriented Mobility Assessment Gait part, RFG: rate of force generation, SPPB: Short Physical Performance Battery, TUG : Timed Up and Go test usual gait speed, TUG FAST : Timed Up and Go test speed as fast as possible without running, TUG 8ft FAST : Timed Up and Go test on 8 feet distance speed as fast as possible, TUG COG FAST : Timed Up and Go test cognitive speed as fast as possible TUG MAN FAST : Timed Up and Go test manual speed as fast as possible, Overall rating: + sufficient, - insufficient, ? indeterminate

** : authors provided additional information (unpublished data)

Where statistical pooling was not possible due to the lack of more than two articles on the same test, the findings of the different tests are presented as extracted from the articles for the specific setting as follows: community-dwelling setting reliability (Table 2) and validity (Table 3), nursing home persons reliability (Table 4) and validity (Table 5), hospital setting reliability (Table 6) and validity (Table 7), and combined setting reliability (Table 8) and validity (Table 9). Supplementary Table 2 shows all cut-offs that were presented in the respective study. Using the optimal cut-off value, the values for SE, SP, LR+, LR-, and post-test probabilities for a positive or a negative test are presented in Tables 3, 5, 7, and 9. When a study could not be pooled and presented several cut-off values, the cut-off presented in the summary of findings tables (Tables 3, 5, 7, and 9) is the one closest to the other included studies that described the same test. To visualize changes for the performance tests for different fall risk prevalence and cut-off values, we created a figure to illustrate the relationship between pre-test and post-test probabilities for different cut-off values (Figure 2).

RESULTS

A total of 6,392 articles were found using our search strategy (PubMed search strategy, supplementary material 3). Of those, 559 duplicates were removed and title and abstract screening was performed for 5,833 articles. This led to excluding an additional 5,415 articles. After a full text reading of the remaining 418 articles, 42 articles were retained for the systematic review, of which 21 articles contained information of reliability and 25 of validity (Figure3: PRISMA flow chart). Four of the included articles contained information for reliability and validity of the included test.

In the following sections, results are ordered according to the chosen settings (community dwelling, nursing home, hospital, and combined setting of hospitalized and community-dwelling persons) in which we found information on either reliability or validity (or both) of performance-based fall risk assessments.

Community-dwelling setting

For the community-dwelling setting, 28 different performance-based measures for the reliability (Table 2), and 16 different measures allowing the calculations of the diagnostic accuracy and post-test probability (Table 3), were included in this systematic review. According to the COSMIN guidelines for selecting outcome measurement instruments, we judged intra-rater and inter-rater reliability with ICC_{agreement} of >0.7 as good (acceptable) values for clinical use (42).

Only one of the 20 tests for which we found information on inter-rater reliability showed ICC_{agreement} values < 0.7, which is not acceptable for clinical use

in a community-dwelling setting (strength assessment with a dynamometer of the dorsiflexion of the foot in sitting position, ICC_{agreement} 0.55 [95% CI 0.11 to 0.81], Table 2). All included assessments for intra-rater reliability showed good ICC_{agreements} > 0.7, acceptable for clinical use. Slightly less studies were performed to assess inter-rater than intra-rater reliability.

Regarding the validity (Table 3), 16 tests could be included. Three of 16 were supported by three or more articles and data could be pooled (BBS (43-47), TUG (43, 47-56), and TUG FAST (57-59)). Seven tests were supported by two articles and six tests were supported by a single study and therefore could not be pooled. BBS was used in five articles with a total sample of 955 participants (43-47). The relatively small changes from pre-test probability to post-test probability for the BBS are presented in Table 3. Eleven studies included the TUG test with a total sample of 3,148 participants (43, 47-56). The TUG FAST test was used in three studies with a total of 269 participants (57-59). Figure 2 identifies the changes from pre-test to post-test probabilities of the TUG FAST test for varying prevalence and cut-off values.

The following values are shown in Table 3. Only seven tests achieved both high values for sensitivity and specificity (BEST (44), BriefBEST(44), MiniBEST (47), TUG FAST pooled (57-59), TUG eight feet (8ft) FAST(58), cognitive TUG FAST (TUG COG FAST) (59) and manual TUG FAST (TUG MAN FAST) (59)). As can be seen, there was a slight tendency of more tests showing lower SE values than SP, indicating that the tests are slightly better to rule in persons who are at risk than to exclude persons who are not at risk. However, the differences between SE and SP are for most tests not very high. Relatively high LR+s were found for TUG COG FAST (59), POMA BALANCE (56, 60) and FRT (60). Relatively low LR-s were found in TUG COG FAST (59), TUG FAST pooled (57-59), and MiniBEST (47). High post-test probabilities for a positive test were found for TUG COG FAST (59), TUG FAST pooled (57-59), POMA BALANCE (56, 60), and FRT (60). Low post-test probabilities for a negative test were found for BEST (44), BriefBEST (44), MiniBEST (44, 47), TUG COG FAST (59), TUG FAST pooled (57-59), FRT (60), and GAIT SPEED FAST (61). The AUC of the tests used in the community-dwelling setting were >0.7 for BEST, MiniBest, BriefBEST, and GAIT SPEED FAST. All other tests showed AUCs <0.7 (Table 1).

Table 2 : Inter- rater and intra-rater reliability for community-dwelling setting, summary of findings

Test	Inter-rater reliability			Intra-rater reliability			Overall rating (+, ?, -)
	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	
BBS	Marques 2016 (44)	28 (1)	0.93 (0.85 to 0.97)	Marques 2016 (44)	28 (1)	0.82 (0.64 to 0.91)	+
BEST	Marques 2016 (44)	60 (2)	0.94 (0.79 to 0.98)	Anson 2017 (74)	165 (4)	0.83 (0.63 to 0.93)	+
	Wang-Hsu 2018(73)			Marques 2016 (44)			
BRIEF BEST	Marques 2016 (44)	28 (1)	0.71 (0.46 to 0.86)	Yingyongyudha 2016 (47)	28 (1)	0.73 (0.49 to 0.87)	+
MINI BEST	Marques 2016 (44)	28 (1)	0.77 (0.56 to 0.89)	Anson 2017 (74)	95 (3)	0.83 (0.73 to 0.90)	+
				Marques 2016 (44)			
TUG	/	/	/	Yingyongyudha 2016 (47)	12 (1)	0.92 (0.73 to 0.98)	+
TUG FAST	Shumway-Cook 2000 (59)	30 (1)	0.98 (0.96 to 0.99)	Hofheinz 2010 (75)	50 (3)	0.93 (0.87 to 0.97)	+
				Lee 2016 (76)			
TUG FAST 6m	/	/	/	Smith 2016 (77)	15 (1)	0.95 (0.85 to 0.98)	+
TUG FAST 9m	/	/	/	Lee 2016 (76)	15 (1)	0.96 (0.88 to 0.99)	+

Table 2 Continued.

Test	Inter-rater reliability			Intra-rater reliability			Overall rating (+, ?, -)
	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	
TUG COG FAST	Marques 2016 (44)	28 (1)	0.99 (0.98 to 1.0)	Hofheinz 2010 (75)	35 (2)	0.97 (0.88 to 0.99)	+
TUG MAN FAST	Marques 2016 (44)	28 (1)	0.99 (0.98 to 1.0)	Smith 2016 (77)			
				Hofheinz 2010 (75)	35 (2)	0.97 (0.81 to 1.0)	+
GAIT SPEED	Hartmann 2009 (78)	23 (1)	0.94 (0.86 to 0.97)	Smith 2016 (77)			
				Goldberg 2011 (79)	113 (3)	0.93 (0.82 to 0.97)	+
GAIT SPEED FAST	/	/	/	Hars 2013 (80)			
				Hartmann 2009 (78)	60 (1)	0.9 (0.84 to 0.94)	+
GAIT SPEED FOAM	Hartmann 2009 (78)	23 (1)	0.96 (0.91 to 0.98)	Hartmann 2009 (78)	23 (1)	0.98 (0.95 to 0.99)	+
GAIT SPEED FOAM DUAL TASK	Hartmann 2009 (78)	23 (1)	0.95 (0.88 to 0.98)	Hartmann 2009 (78)	23 (1)	0.93 (0.84 to 0.97)	+
DYNAMOMETER HIP FLEX SITTING	Arnold 2010 (81)	18 (1)	0.84 (0.62 to 0.94)	Arnold 2010 (81)	43 (2)	0.90 (0.81 to 0.95)	+
				Ford-Smith 2001 (82)			
DYNAMOMETER HIP FLEX STANDING	Arnold 2010 (81)	18 (1)	0.85 (0.64 to 0.94)	Arnold 2010 (81)	18 (1)	0.88 (0.69 to 0.95)	+
DYNAMOMETER HIP EXT STANDING	Arnold 2010 (81)	18 (1)	0.92 (0.81 to 0.97)	Arnold 2010 (81)	18 (1)	0.83 (0.59 to 0.93)	+

Table 2 Continued.

Inter-rater reliability				Intra-rater reliability			
Test	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Overall rating (+, ?, -)
DYNAMOMETER HIP EXT SUPINE	/	18 (1)	/	Ford-Smith 2001 (82)	25 (1)	0.74 (0.49 to 0.88)	+
DYNAMOMETER HIP ABD SUPINE	Arnold 2010 (81)	18 (1)	0.84 (0.62 to 0.94)	Arnold 2010 (81)	18 (1)	0.89 (0.72 to 0.96)	+
DYNAMOMETER HIP ABD STANDING	Arnold 2010 (81)	18 (1)	0.92 (0.78 to 0.97)	Arnold 2010 (81)	18 (1)	0.94 (0.83 to 0.98)	+
DYNAMOMETER KNEE FLEX				Ford-Smith 2001 (82)	25 (1)	0.85 (0.68 to 0.93)	+
DYNAMOMETER KNEE EXT 45°	Arnold 2010 (81)	18 (1)	0.82 (0.58 to 0.93)	Arnold 2010 (81)	18 (1)	0.86 (0.66 to 0.95)	+
DYNAMOMETER KNEE EXT 90°	/	/	/	Ford-Smith 2001 (82)	211 (2)	0.94 (0.88 to 0.97)	+
DYNAMOMETER PLANTARFLEX	/	/	/	Katoh 2014 (83)	25 (1) ¹	0.71 (0.44 to 0.86)	+
DYNAMOMETER DORSIFLEX SITTING	Arnold 2010 (81)	18 (1)	0.55 (0.11 to 0.81)	Arnold 2010 (81)	18 (1) ¹	0.80 (0.53 to 0.92)	+
DYNAMOMETER DORSIFLEX LYING	/	/	/	Ford-Smith 2001 (82)	25 (1)	0.84 (0.67 to 0.93)	+

Table 2 Continued.

Inter-rater reliability				Intra-rater reliability			
Test	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Overall rating (+, ?, -)
CALF RAISE SENIOR	André 2016 (84)	12 (1)	0.84 (0.51 to 0.95)	André 2016 (84)	41 (1)	0.90 (0.82 to 0.95)	+

ABD : Abduction, BBS : Berg Balance Scale, BEST : Balance Evaluation System Test, BriefBEST : Brief Balance Evaluation System Test, MiniBEST : Mini Balance Evaluation System Test, 95% CI : 95% confidence interval, DYNAMOMETER : strength measures with dynamometer in the respective position, EXT : Extension, FLEX : Flexion, Gait speed : Gait speed at usual speed, Gait Speed FAST : Gait speed as fast as possible without running, ICC_{agreement} : Intraclass correlation coefficient agreement, TUG : Timed Up and Go test usual gait speed, TUG FAST : Timed Up and Go test speed as fast as possible without running, TUG FAST 6m : Timed Up and Go test speed as fast as possible without running on 6m distance, TUG FAST 9m : Timed Up and Go test speed as fast as possible without running on 9m distance, TUG COG FAST : Timed Up and Go test cognitive; speed as fast as possible TUG MAN FAST : Timed Up and Go test manual; speed as fast as possibleOverall rating: + sufficient, - insufficient, ? indeterminate, / : no values

Table 3 : Validity for community-dwelling setting, summary of findings

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95 CI)	Specificity (95% CI)	LR+	LR- probability +test	Post-test probability % -test
BBS	Pooled	949 (5)	30% (2)	≤ 54.13 points	0.65 (0.25-0.91)	0.52 (0.21- 0.81)	1.34	0.68	37 23
	Greene 2010 (43)								
	Marques 2016 (44)								
	Muir 2008 (45)								
	Santos 2011 (46)								
	Yingyongyudha (2016) (47)								
BEST	Marques (2016) (44)	122 (1)	30%	48.5* points	0.74 (-)	0.72 (-)	2.59	0.37	53 13
BEST	Yingyongyudha (2016) (47)	200 (1)	30%	66* points	0.76 (0.67-0.83)	0.5 (0.40-0.60)	1.52	0.48	39 17
BRIEF BEST	Marques (2016) (44)	122 (1)	30%	12.5* points	0.74 (-)	0.71 (-)	2.49	0.38	52 14
MINI BEST	Marques (2016) (44)	122 (1)	30%	19.5* points	0.74 (-)	0.67 (-)	2.23	0.40	49 14
MINI BEST	Yingyongyudha (2016) (47)	200 (1)	30%	16* points	0.85 (0.77-0.9)	0.75 (0.66-0.83)	3.40	0.20	59 8
POMA	Murphy 2003 (60)	45 (1)	30%	12* points	0.55 (-)	0.97 (-)	18.33	0.46	89 16
BALANCE									
POMA	Trueblood 2001 (56)	180 (1)	30%	12* points	0.21 (-)	0.95 (-)	4.20	0.83	64 26
BALANCE									
POMA GAIT	Trueblood 2001 (56)	180 (1)	30%	8* points	0.24 (-)	0.91 (-)	2.67	0.84	53 26

Table 3 Continued.

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95 CI)	Specificity (95% CI)	LR+	LR- probability +test	Post-test probability % -test
TUG	Pooled	3148 (11)	30%	≥13.20 s	0.39 (0.22-0.59)	0.74 (0.63-0.88)	1.72	0.79	43 25
	Abu Samah 2018 (48)								
	Alexandre 2012 (49)								
	Greene 2010 (43)								
	Trueblood 2001 (56)								
	Chantanachai 2014 (50)								
	Criter 2016 (51)								
	Criter 2017 (52)								
	Ibrahim. 2017 (53)								
	Kojima 2015 (54)								
	Shimada 2009 (55)								
	Yingyongyudha (2016) (47)								
Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95 CI)	Specificity (95% CI)	LR+	LR- probability +test	Post-test probability % -test
TUG FAST	Pooled	269 (3)	30%	≥5.6s	0.86 (0.49- 0.97)	0.76 (0.05- 1.00)	3.64	0.19	61 7
	Arnold 2007 (57)								
	Rose 2002 (58)								
	Shumway-Cook 2000 (59)								

Table 3 Continued.

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95 CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability % +test	Post-test probability % -test
TUG 8feet FAST	Rose 2002 (58)	134 (1)	30%	≥8.5s	0.78 (-)	0.86 (-)	5.57	0.26	70	10
TUG COG FAST	Hofheinz 2016 (85)	113 (1)	30%	≤ 10.3s **	0.49 (0.33-0.65)	0.77 (0.68-0.86)	2.13	0.66	48	22
TUG COG FAST	Shumway-Cook 2000 (59)	30 (1)	30%	≥14.5s	0.86 (-)	0.93 (-)	12.90	0.15	84	6
TUG MAN FAST	Shumway-Cook 2000 (59)	25 (1)	30%	≥15s	0.80	0.93	12.00	0.21	82	9
FRT	Murphy 2003 (60)	45 (1)	30%	20.32cm*	0.73 (-)	0.88 (-)	6.08	0.31	72	12
FRT	Shimada 2009 (55)	455 (1)	30%	≤ 18cm	0.47 (-)	0.59 (-)	1.15	0.90	33	28
GAIT SPEED	Middleton 2016 (61)	217 (1)	30%	< 0.76 m/s	0.65 (-)	0.71 (-)	2.25	0.49	49	17
GAIT SPEED	Shimada 2009 (55)	455 (1)	30%	≤ 0.7 m/s	0.56 (-)	0.59 (-)	1.37	0.75	37	24
GAIT SPEED FAST	Middleton 2016 (61)	217 (1)	30%	< 1.13 m/s	0.77 (-)	0.60 (-)	1.92	0.39	45	14
GAIT SPEED FAST	Shimada 2009 (55)	455 (1)	30%	≤ 1 m/s	0.58 (-)	0.58 (-)	1.38	0.72	37	24
GRIP	Shimada 2009 (55)	455 (1)	30%	≤17 kg	0.55 (-)	0.55 (-)	1.22	0.82	34	26

Table 3 Continued.

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95 CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability % +test	Post-test probability % -test
OLS	Shimada 2009 (55)	455 (1)	30%	≤3 s	0.51 (-)	0.61 (-)	1.31	0.80	36	26

BBS : Berg Balance Scale, BEST : Balance Evaluation System Test, BriefBEST : Brief Balance Evaluation System Test, MiniBEST : Mini Balance Evaluation System Test, CI : 95% confidence interval, FRT : Functional Reach Test, Gait speed : Gait speed at usual speed, Gait Speed FAST : Gait speed as fast as possible without running, GRIP : grip strength, LR+ : likelihood ratio for a positive test, LR- : likelihood ratio for a negative test, OLS : One Leg Stance, POMA Balance : Performance Oriented Mobility Assessment Balance part, POMA GAIT : Performance Oriented Mobility Assessment Gait part, TUG : Timed Up and Go test usual gait speed, TUG FAST : Timed Up and Go test speed as fast as possible without running, TUG 8ft FAST : Timed Up and Go test on 8 feet distance speed as fast as possible, TUG COG FAST : Timed Up and Go test cognitive speed as fast as possible TUG MAN FAST : Timed Up and Go test manual speed as fast as possible, +test : positive test, -test : negative test (-) : confidence interval not reported

* Not stated if ≥, ≤
** cut-off closest to the other articles included for the same test. In the original article four other cut-off values are presented (see supplementary Table 2)

Nursing home setting

For nursing home patients, five assessments were identified for both reliability and validity. All five assessments found for inter-rater and intra-rater reliability showed a good ICC_{agreement} and in terms of reliability, can be recommended for clinical use (Table 4). There were not enough articles available to pool information for the tests for diagnostic accuracy. The following values for diagnostic accuracy are in Table 5. MiniBEST (62) was the only test that showed relatively high values for both SE and SP. TUG (63) was the only test showing a high SE compared to SP, and all other tests showed higher SP than SE. High LR+s were found for BBS (62) and BriefBEST (62). Low LR-s were found for TUG. Regarding post-test probability for a positive test, BBS (62), BEST (62), and BriefBEST (62) showed high values. Only TUG (63) showed a low post-test probability for negative tests. The AUCs of five out of eight tests could be calculated. Only the AUC of the TUG was < 0.7 (Table 1).

Hospital setting

Three tests were identified for reliability (Table 6) and for validity (Table 7) for hospitalized patients. All three assessments (variations of gait speed assessments) can be recommended for clinical use regarding their intra-rater reliability (ICC_{agreement} >0.7) (Table 6). None of the included studies examined inter-rater reliability. Three out of four studies showed higher SE than SP (simplified POMA (32), SPPB (32), and TUG (64), Table 7). All LR+ and LR- values were close to one (Table 7). The post-test probabilities of a positive test were relatively low and the post-test probabilities for a negative test were all relatively high (Table 7). The AUCs of all tests were < 0.7 (Table 1).

Combined setting of hospitalized and community-dwelling persons

For the combined setting of hospitalized and community-dwelling persons, eight tests for reliability (Table 8) and thirteen for validity (Table 9) were included in this systematic review. Two assessments were included for inter-rater reliability and showed good ICC_{agreement} (>0.7). For intra-rater reliability, eight different measures were included and showed good ICC_{agreement} (> 0.7) as well (Table 8).

Regarding validity (Table 9), all tests showed high sensitivity > 0.9 and low specificity < 0.5. LR+s are close to one for all included tests. LR-s were low for all the following tests: (HIP STRENGTH ABD, HIP STRENGTH ABD RFG, HIP STRENGTH EXTERNAL ROTATORS, HIP STRENGTH INTERNAL ROTATORS, and INTERNAL ROTATORS RFG) (65). The post-test probabilities of a positive test are all relatively

low. All tests, except two (HIP STRENGTH ADDUCTORS and HIP STRENGTH ADDUCTORS RFG), showed low post-test probabilities after a negative test. Six out of 15 tests (HIP STRENGTH ABDUCTORS, HIP STRENGTH ABDUCTORS RFG, HIP STRENGTH ADDUCTORS, HIP STRENGTH ADDUCTORS RFG, HIP STRENGTH EXTERNAL ROTATOR, AND HIP STRENGTH FLEXORS) showed an AUC > 0.7 (Table 1).

Quality assessment

Using COSMIN RoB to assess the reliability studies' methodological quality showed that only one study was rated very good, all other studies showed, at most, adequate; however, most studies resulted in doubtful or inadequate risk of bias (Supplementary Table 1, reliability). The overall quality assessment of the articles with the QUADAS-2 tool (Figure 4) included for validity showed that for the domains of patient selection, index test, and reference standard, about 10% to 20% showed a low risk of bias and about 60% to 80% of the included articles showed an unclear risk of bias. A high risk of bias was detected for 5% to 20% of the studies for the domains of patient selection and index test. For the domain of flow and timing, about 95% of the included articles showed a low risk of bias. Almost all articles (about 95%) showed a low risk of bias for applicability (detailed rating of COSMIN Risk of bias tool [Supplementary Table 1]) and QUADAS-2 (supplementary Table 2).

The overall rating based on the COSMIN group's criteria for good measurement properties (17) showed that for reliability, only one test in the community-dwelling setting was rated "insufficient" (DYNAMOMETER DORSIFLEX SITTING). All other tests over all the chosen settings reached a "sufficient" rating. For validity, the criteria for good measurement properties are judged via the AUC values in Table 1. The ratings for every setting list that about half of the tests had "sufficient" values and the other half had "insufficient" values. AUCs could not be calculated for all tests based on the data published in the respective articles. However, if confidence intervals were included in the overall rating for good measurement properties of the validity studies, only HIP STRENGTH ABDUCTORS ended slightly above 0.7 and would thus, keep the rating "sufficient ." If confidence intervals were included in the rating for reliability, for the community-dwelling setting, about half of the tests would change from a "sufficient" to an "insufficient" rating. The same is true for the nursing home population's three tests. No changes in the rating would occur for the hospital and the combined setting.

Table 4 : Inter-rater and intra-rater reliability for nursing-home setting, summary of findings

Inter-rater reliability				Intra-rater reliability			
Test	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Overall rating (+, ?, -)
BBS	Holbein–Jenny 2005 (86)	76 (2)	0.97 (0.61 to 1.00)	Holbein–Jenny 2005 (86)	64 (2)	0.84 (0.69 to 0.92)	+
	Viveiro 2018 (62)			Viveiro 2018 (62)			
BEST	Viveiro 2018 (62)	49 (1)	0.99 (0.99 to 1.00)	Viveiro 2018 (62)	37 (1)	0.94 (0.89 to 0.97)	+
MiniBEST	Viveiro 2018 (62)	49 (1)	0.99 (0.99 to 1.00)	Viveiro 2018 (62)	37 (1)	0.93 (0.87 to 0.97)	+
BriefBEST	Viveiro 2018 (62)	49 (1)	0.99 (0.99 to 1.00)	Viveiro 2018 (62)	37 (1)	0.94 (0.88 to 0.97)	+
FRT	Holbein–Jenny 2005 (86)	27 (1)	0.98 (0.96 to 0.99)	Holbein–Jenny 2005 (86)	37 (1)	0.75 (0.52 to 0.88)	+

BBS : Berg Balance Scale, BEST : Balance Evaluation System Test, BriefBEST : Brief Balance Evaluation System Test, MiniBEST : Mini Balance Evaluation System Test, , CI: 95% confidence interval, FRT : Functional Reach Test, ICC: intraclass correlation coefficient agreement, Overall rating: + sufficient, - insufficient, ? indeterminate

Table 5 : Validity for nursing home setting, summary of findings

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR-	Post-test probability %, +test	Post-test probability %, -test
BBS	Viveiro 2018 (62)	49 (1)	40% (2)	48* points	0.55 (0.36-0.55)	0.94 (0.73-1.00)	9.87	0.48	87	24
BEST	Viveiro 2018 (62)	49 (1)	40%	57* points	0.61 (0.42-0.78)	0.83 (0.59-0.96)	3.68	0.46	71	23
BriefBEST	Viveiro 2018 (62)	49 (1)	40%	10* points	0.58 (0.39-0.75)	0.94 (0.73-1.00)	10.45	0.44	87	23

Table 5 Continued.

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR-	Post-test probability %, +test	Post-test probability %, -test
MiniBEST	Viveiro 2018 (62)	49 (1)	40%	10* points	0.71 (0.52-0.86)	0.78 (0.52-0.94)	3.19	0.37	68	20
TUG	Applebaum 2017 (63)	53 (1)	40%	14.2s*	0.97 (-)	0.15 (-)	1.14	0.2	43	12

BBS : Berg Balance Scale, BEST : Balance Evaluation System, BriefBEST : Brief Balance Evaluation System, MiniBEST : Mini Balance Evaluation System, 95% CI : 95% confidence interval, s : second, +test : positive test, -test : negative test, TUG: Timed Up and Go, (-) : confidence interval not reported
* Not stated if $\geq$, $\leq$

Table 6 : Inter-rater and intra-rater reliability for hospital setting, summary of findings

Inter-rater reliability				Intra-rater reliability			
Test	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Overall rating (+, ?, -)
GAIT SPEED	/	/	/	Hars 2013 (80)	60 (1)	0.92 (0.87 to 0.95)	+
GAIT SPEED FAST	/	/	/	Hars 2013 (80) Martinez 2016 (87)	170 (2)	0.98 (0.87 to 1.0)	+
GAIT SPEED DUAL TASK	/	/	/	Hars 2013 (80)	60 (1)	0.91 (0.85 to 0.95)	+

95% CI : 95% confidence interval, GAIT SPEED : Gait speed usual walking speed ; GAIT SPEED FAST : Gait speed as fast as possible without running, GAIT SPEED DUAL TASK: Gait speed with dual task conditions, ICC_{agreement} : Intraclass correlation coefficient agreement, Overall rating: + sufficient, - insufficient, ? indeterminate, / : no values

Table 7: Validity for hospital setting, summary of findings

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR- probability %	Post-test probability %	Post-test probability %	Post-test probability %
SIMPLIFIED POMA	Hars 2018 (32)	524 (1)	24% (32, 33)	≥ 2 points	0.71 (-) **	0.46 (-)**	1.31	0.64	47	30	30
SPPB	Hars 2018 (32)	678 (1)	24%	< 5 points	0.68 (-)**	0.46 (-)**	1.26	0.69	46	32	32
TUG	Chow 2018 (64)	192 (1)	24%	≥ 12s	0.71 (-)	0.28 (-)	0.98	1.04	40	41	41
TUG	Hars 2018 (32)	525 (1)	24%	> 29.5s	0.41 (-)**	0.69 (-)**	1.34	0.85	47	36	36

95% CI : 95% confidence interval, SimplifiedPOMA : simplified Performance Oriented Mobility Assessment, SPPB : Short Physical Performance Battery, TUG : Timed Up and Go test, s : seconds, (-) : confidence interval not reported

**: authors provided additional information (unpublished data)

Table 8 : Inter-rater and intra-rater reliability for combined setting of hospitalized and community-dwelling persons, summary of findings

Inter-rater reliability					Intra-rater reliability				
Test	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Author	No of participants (n studies)	Relative Reliability ICC _{agreement} (95% CI)	Overall rating (+, ?, -)	Relative Reliability ICC _{agreement} (95% CI)	Overall rating (+, ?, -)
GAIT SPEED	Bautmans 2011 (88)	81 (1)	0.94 (0.90 to 0.96)	Bautmans 2011 (88)	81 (1)	0.96 (0.93 to 0.97)	+	0.96 (0.93 to 0.97)	+
GAIT SPEED MEAN OF TWO	Bautmans 2011 (88)	81 (1)	0.98 (0.96 to 0.98)	Bautmans 2011 (88)	81 (1)	0.99 (0.98 to 1.0)	+	0.99 (0.98 to 1.0)	+
DYNAMOMETER HIP ABD STANDING	/	/	/	Bruyneel 2018 (89)	32 (1)	0.98 (0.96 to 0.99)	+	0.98 (0.96 to 0.99)	+
DYNAMOMETER HIP ABD STANDING RFG	/	/	/	Bruyneel 2018 (89)	32 (1)	0.93 (0.86 to 0.97)	+	0.93 (0.86 to 0.97)	+

Table 8 Continued.

Inter-rater reliability					Intra-rater reliability				
DYNAMOMETER HIP ABD SIDELYING MAX	/	/	/	Gafner 2017 (90)	76 (1)	0.94 (0.91 to 0.96)	+	0.94 (0.91 to 0.96)	+
DYNAMOMETER HIP ABD SIDELYING RFG	/	/	/	Gafner 2017 (90)	76 (1)	0.94 (0.91 to 0.96)	+	0.94 (0.91 to 0.96)	+
DYNAMOMETER HIP ADD SIDELYING MAX	/	/	/	Gafner 2017 (90)	76 (1)	0.90 (0.85 to 0.94)	+	0.90 (0.85 to 0.94)	+
DYNAMOMETER HIP ADD SIDELYING RFG	/	/	/	Gafner 2017 (90)	76 (1)	0.94 (0.91 to 0.96)	+	0.94 (0.91 to 0.96)	+

ABD: Abduction, ADD: Adduction, 95% CI : 95% confidence interval , GAIT SPEED: Gait speed usual walking speed, ICC_{agreement} : Intraclass correlation coefficient agreement, MAX: maximum strength, RFG: Rate of force generation, Overall rating: + sufficient, - insufficient, ? indeterminate, / : no values

Table 9: Validity for combined setting of hospitalized and community-dwelling persons, summary of findings

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR- probability %	Post-test probability %	Post-test probability %	Post-test probability %
Grip	Gafner 2018 (65)	56 (1)	27% (2, 32, 33)*	≤70 kPa	0.91 (-)**	0.32 (-)**	1.3**	0.3**	33**	10**	10**
HIP STRENGTH ABDUCTORS	Gafner 2018 (65)	56 (1)	27%	≤1.1 N/kg	0.91 (-)	0.43 (-)	1.6	0.2	37	8	8
HIP STRENGTH ABDUCTORS RFG	Gafner 2018 (65)	56 (1)	27%	≤ 8.5 N/kg/s	0.91 (-)	0.43 (-)	1.6	0.2	37	8	8
HIP STRENGTH ADDUCTORS	Gafner 2018 (65)	56 (1)	27%	≤ 1.3 N/kg	0.90 (-)**	0.19 (-)**	1.1**	0.5**	29**	17**	17**

Table 9: Continued.

Test	Author	No of participants (n studies)	Pre-test probability	Optimal cut-off	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR-	Post-test probability %, +test	Post-test probability %, -test
HIP STRENGTH ADDUCTORS RFG	Gafner 2018	(65) 56 (1)	27%	≤ 6.2 N/kg/s	0.90 (-)**	0.15 (-)**	1.1**	0.7**	28**	20**
HIP STRENGTH EXTENSORS	Gafner 2018	(65) 56 (1)	27%	≤ 1.7 N/kg	0.90 (-)**	0.32 (-)**	1.3**	0.3**	33**	10**
HIP STRENGTH EXTENSORS RFG	Gafner 2018	(65) 56 (1)	27%	≤ 10.7 N/kg/s	0.90 (-)**	0.18 (-)**	1.1**	0.6**	29**	17**
HIP STRENGTH EX.ROTATORS	Gafner 2018	(65) 56 (1)	27%	≤ 0.9 N/kg	0.91 (-)**	0.50 (-)**	1.8**	0.2**	40**	7**
HIP STRENGTH EX. ROTATORS RFG	Gafner 2018	(65) 56 (1)	27%	≤ 5.7 N/kg/s	0.91 (-)**	0.32 (-)**	1.3**	0.3**	33**	10**
HIP STRENGTH FLEXORS	Gafner 2018	(65) 56 (1)	27%	≤ 1.9 N/kg	0.90 (-)**	0.25 (-)**	1.2**	0.4**	31**	13**
HIP STRENGTH FLEXORS RFG	Gafner 2018	(65) 56 (1)	27%	≤ 14.5 N/kg/s	0.90 (-)**	0.25 (-)**	1.2**	0.4**	31**	13**
HIP STRENGTH IN. ROTATORS	Gafner 2018	(65) 56 (1)	27%	≤ 1.4 N/kg	0.91 (-)**	0.43 (-)**	1.6**	0.2**	37**	8**
HIP STRENGTH IN. ROTATORS RFG	Gafner 2018	(65) 56 (1)	27%	≤ 6.4 N/kg/s	0.91 (-)**	0.43 (-)**	1.6**	0.2**	37**	8**

95% CI: 95% confidence interval, EX.ROTATORS: external rotators, Grip: Grip strength, RFG: rate of force generation, IN.ROTATORS: internal rotators; kPa: kilo pascal, LR+: Likelihood ratio for a positive test, LR-: Likelihood ratio for a negative test, N/kg: Newton per kilo, (-) : confidence interval not reported

* Chosen prevalence between the fall risk prevalence of hospitalized and community-dwelling persons

** : authors provided additional information (unpublished data)

DISCUSSION

This systematic review aimed to provide an overview of the reliability and diagnostic accuracy of commonly used performance tests to assess older persons' fall risk in more relevant settings. We desire for the results of this study to improve the decision-making process concerning an optimal choice for suitable single- or multidimensional fall risk assessments for older persons in specific settings with varying fall risk prevalence. Additionally, we want to support the individual decision-making process between patient and caregiver.

For the community-dwelling setting, results indicate that all except one test (DYNAMOMETER DORSIFLEX SITTING) can be recommended for clinical use. They fulfilled the prerequisite of good reliability and allowed good discriminative validity. Taking into consideration the sensitivity and the post-test probability of a negative test, authors can recommend the MiniBEST (47), TUG FAST pooled (57-59), and TUG COG FAST (59) for the multidimensional fall risk assessments, and the FRT (60), and GAIT SPEED FAST (61) as single-performance tests.

For nursing home patients, only multidimensional fall risk assessments were studied, all of which showed acceptable reliability for clinical use. For discriminative validity, the TUG might be recommended as it showed high sensitivity and low post-test probability for a negative test in one study (63).

For the hospital setting, all included tests (single-performance tests) showed high reliability, but none of the included multidimensional fall risk assessments for the discriminative validity can be recommended.

For a mixed setting of hospitalized and community-dwelling persons, all tests can be recommended with regard to their reliability. In regard to their discriminative ability, tests of the lower limb strength (HIP STRENGTH EXTERNAL ROTATORS, HIP STRENGTH ABDUCTORS, HIP STRENGTH ABDUCTORS RFG, HIP STRENGTH INTERNAL ROTATORS AND HIP STRENGTH INTERNAL ROTATORS RFG) and hand-grip strength are recommended.

However, these recommendations have limited strength and should be applied with caution, as we cannot be completely certain about the tests' results. While conducting this review, it became clear that there were three main reasons limiting us from giving strong recommendations: a) most of the included studies, even when pooled (TUG FAST), had small sample sizes; b) for some settings, only a very limited amount of tests and studies could be included; and c) good reliability was a prerequisite for good criterion validity; however, we could not find for all tests, information of both reliability (inter- and intra-rater) and criterion validity for the same setting.

We saw an imbalance in the number of investigated multidimensional and single-performance tests for all settings except the community-dwelling setting in which they were equally distributed.

For the nursing home setting, the FRT was the only available single-performance measure, and only data on reliability was presented. All other investigated tests (five) were multidimensional performance measures.

For the hospital setting tests, only the reliability of single-performance measures (three) was investigated, whereas the validity was only assessed for multidimensional measures (three).

On the other hand, for the combined setting of hospitalized and community-dwelling persons, exclusively single-performance tests were found for reliability and validity.

It seems surprising that single-performance tests were less investigated as they usually save time and strongly indicate what should be targeted during treatment. The diagnostic accuracy of single-performance tests should be investigated more and with an improved methodological approach to allow their wider daily use in clinics. In all settings, multidimensional fall risk assessments showed a tendency of higher values for sensitivity and lower post-test probabilities than single-performance measures and thus seem to be better for ruling out fallers. For the combined setting of hospitalized and community-dwelling persons, only single-performance tests could be found and a direct comparison to multidimensional fall risk tests is not possible. The higher number of studies in community-dwelling participants might indicate that a) the chosen performance-based measures are not feasible for hospitalized or nursing home patients, and b) not enough studies are performed for this frail population.

As already highlighted in the introduction, falls are frequent in persons aged 65 years and older (1, 2) and can have serious consequences for an affected person (3-5), making fall risk detection and prevention a global public health challenge and thus a matter of priority. However, our systematic review showed that the included tests' discriminative ability to detect persons at risk of falling was in general not very good for clinical use. Most of the included tests only showed relatively low sensitivity values and small changes from pre- to post-test probabilities.

In this systematic review, we focused on performance-based measures. However, falls have a multifactorial etiology that can be categorized in environmental factors, psychological factors, cognitive factors, factors linked to medication, and physiological factors (66). Our results, as well as the knowledge of the multifactorial etiology of falls, underline that the currently used fall risk assessments are not good enough to discriminate between older fallers and non-fallers when used alone. Lusardi et al.'s (2017) systematic review confirms findings (6). Combined use of

measurement instruments will allow the detection of specific components that have to be targeted for an effective prevention and/or rehabilitation program.

One of this systematic review's large advantages is its clinical applicability. The presentation of the reliability, together with the diagnostic accuracy of fall risk measures, is important for clinical use. The explanation and presentation of post-test probabilities and likelihood ratios emphasize the focus on the clinical application of this study's results, but also facilitates the understanding of applying the results of other measures investigated in other articles. Our extensive search strategy allowed us to detect a large number of articles. The presentation of our results separately for the different settings is a further strength of this systematic review. As fall risk prevalence influences the diagnostic accuracy of measures, separating the settings is a necessary step and should not be neglected. Additionally, it facilitates clinicians in choosing the most optimal measure for their specific patient and personal environment. From our literature search, we also identified studies with a combined setting of hospitalized and community-dwelling persons. Two points must be mentioned for this combined setting: a) not many articles could be included, and b) the spectrum bias can influence the results of diagnostic accuracy studies and might lead to its overestimation (19). The analyses' results within the combined setting of hospitalized and community-dwelling patients might have been influenced by this bias and should be handled accordingly.

We faced several challenges in summarizing and interpreting our findings during this systematic review. The combination of articles using fall history and the prospective number of falls as a reference standard was used as best-evidence synthesis in this systematic review (6). However, using data from the number of falls within a defined period in the past as a reference standard, could be vulnerable to recall bias (67). Additionally, because falls are likely to be underreported by older adults (68, 69), this could influence the diagnostic accuracy by participants' potential misclassification. Preferably, falls are assessed in a prospective manner with close monitoring of the participants' fall rate. However, most of the included studies used fall history (≥ 6 months) as a reference standard, and the best-evidence synthesis combining studies with the reference standards fall history of 6 months or more, and others with a fall history of 12 months or more, might influence the results. Additionally, there are different classifications of fallers versus non-fallers in the current literature. Some studies count persons with one or no falls as non-fallers and those with two or more falls as fallers (70). Others classified only people with no falls in the non-faller group and persons with one or more falls as fallers (32, 62). This systematic review used the latter categorization, thus some articles had to be excluded due to another definition of the faller versus non-fallers group. Furthermore, due to the exclusion criteria of fall-influencing diseases, we unfortunately also had

to exclude different studies of the developers of some of the measures as they often included a population with various diseases. Further, the COSMIN group is rating, in their updated criteria, ICCs and AUCs ≥ 0.7 as a good measurement property (17). Following these criteria, the results of this systematic review were classified as sufficient, insufficient, or indeterminate, and thus we could for example recommend all but one test as reliable for clinical use. Nonetheless, if the confidence intervals were considered for decision making, eight out of 19 tests for inter-rater reliability and 11 out of 28 tests for intra-rater reliability could not be recommended for clinical use in the community-dwelling setting.

Evaluating the cut-off values presented some challenges as well. Different methodologies were used to calculate cut-off values to evaluate diagnostic accuracy. Most authors used the Youden Index, which gives equal weight to SE and SP (71), to calculate the cut-off values. Using the Youden Index implies that the “costs” of false-positive and false-negative results are equivalent (71). However, this is not often the case in daily clinical life, thus other methods exist for their calculation. For example, a certain sensitivity and/or specificity level to determine cut-off values can be decided.

Another point to address is the vast number of different cut-off values that exist. As the diagnostic accuracy of the included measures is not very high, the question remains if a dichotomization of a test result is clinically appropriate or if the test results should rather be used as a probability of fall risk. The decision to categorize older persons into the group of persons at risk of falling (fallers) or persons who are not at risk of falling (non-fallers) should be a joint interdisciplinary process of health care professionals and the concerned persons, considering the results of several fall risk assessment tests, the persons’ settings, the concerned persons’ personal goals, and reported additional information. For example, the Stopping Elderly Accidents, Deaths, and Injuries (STEADI) fall prevention tool kit suggests asking the participant to answer some basic questions (about the fall history, fear of falling, unsteadiness while standing or walking), and perform three basic fall risk assessments adapted for the patient’s specific setting (eg TUG FAST, miniBEST, and FRT) for a general assessment of the fall risk (72). As the assessment of fall risk alone does not decrease falls, a specific assessment of different fall-influencing factors should be performed to obtain targeted information about the specific fall risk factors to treat (72). To be effective, health care professionals should provide individualized intervention, together with supplemented educational material, following the fall risk assessment (72).

CONCLUSION

This review summarizes performance-based measures for detecting fall risk in older persons. A clinical application of its results in everyday practice can easily be done on both a population and individual level, and supports the decision-making process between patient and caregiver.

To choose an appropriate fall risk assessment, health care professionals should know and understand a test’s clinimetric properties. The prevalence of fall risk influences the diagnostic accuracy of the measurement tools and is known to vary between different settings. It was highly important to present the results of the four relevant settings separately. Different results were found for each setting. The most extensive information was found for community-dwelling persons. However, in the other settings, only a limited amount of studies could be included. For all settings, the sample sizes of most of the included studies were rather small. Therefore, it is difficult to present firm conclusions. Nonetheless, for each setting, we made a separate recommendation based on the outcomes regarding reliability and diagnostic accuracy. The relevance of working with likelihood ratios and how to use the Fagan nomogram to obtain information on post-test probabilities is underlined for clinical practice.

ACKNOWLEDGEMENTS

The authors would like to thank Gregor Franssen, librarian at the University of Maastricht, who helped develop the search strategy and approved its final version.

FUNDING

This study was funded by the internal grant of the HES-SO University of Applied Sciences and Arts of Western Switzerland. The sponsor had no role in the conception, design, methods, data analysis, and preparation of this article.

DISCLOSURE

The authors declare no conflicts of interest in this work.

SUPPLEMENTARY MATERIAL

Supplementary Material: Tables 1 and 2

The detailed data extraction tables (Table 1 for reliability, Tables 2 for validity) are available as additional Excel file (named “Supplementary_Material_Table1_and_Table2”) to this dissertation.

<https://www.dovepress.com/reliability-and-diagnostic-accuracy-of-commonly-used-performance-tests-peer-reviewed-fulltext-article-CIA>

Supplementary Material 3: Detailed PubMed search strategy

((((((((((((((((((accidental falls[MeSH Terms]) OR fall[tiab]) OR falls[tiab]) OR falling[tiab])
OR faller[tiab]) OR stumble[tiab]) OR stumbling[tiab]) OR trip[tiab]) OR tripping[tiab])
OR slip[tiab]) OR slipping[tiab])) AND (((((((((((((((((((«Aged»[Mesh]) OR
 («Aged, 80 and over»[Mesh])) OR «Frail Elderly»[Mesh]) OR «Geriatrics»[Mesh]
OR elder*[tiab]) OR eldest[tiab]) OR frail*[tiab]) OR geriatri*[tiab]) OR old*[tiab]
OR old age*[tiab]) OR oldest old*[tiab]) OR senior*[tiab]) OR senium[tiab]) OR very
old*[tiab]) OR septuagenarian*[tiab]) OR octagenarian*[tiab]) OR octogenarian*[tiab]
OR nonagenarian*[tiab]) OR centarian*[tiab]) OR centenarian*[tiab]) OR
supercentenarian*[tiab]) OR older people[tiab]) OR older subject*[tiab]) OR older
patient*[tiab]) OR older age*[tiab]) OR older adult*[tiab]) OR older man[tiab]) OR
older men[tiab]) OR older male*[tiab]) OR older woman[tiab]) OR older women[tiab]
OR older female*[tiab]) OR older population*[tiab]) OR older person*[tiab])) AND
((((((((((((((((((((timed up[tiab] AND go[tiab])) OR timed upand go[tiab]) OR (time
up[tiab] AND go[tiab])) OR (get up and go[tiab])) OR TUG[tiab]) OR GUG[tiab]) OR
TGUG[tiab]) OR TGUGT[tiab]) OR ETGUG[tiab]) OR TUGT[tiab]) OR TGUGM[tiab]))
OR (((Mini-BESTest[tiab]) OR MiniBESTest[tiab]) OR BESTest[tiab]) OR balance
evaluation system*[tiab])) OR ((berg balance[tiab]) OR BBS[tiab])) OR ((short
physical performance[tiab]) AND SPPB[tiab])) OR (((performance oriented mobility
assessment[tiab]) OR POMA[tiab]) OR tinetti[tiab])) OR (((((((Walking Speed[Mesh]
OR walking speed*[tiab]) OR walking velocit*[tiab]) OR walking pace*[tiab])
OR gait speed*[tiab]) OR gait characteristics[tiab]) OR gait velocit*[tiab]) OR
ambulation speed*[tiab]) OR spatio temporal parameter*[tiab]) OR spatiotemporal
parameter*[tiab])) OR ((((((single leg*[tiab]) OR SLS[tiab]) OR one leg*[tiab])
OR single limb*[tiab]) OR OLS[tiab]) OR OLST[tiab])) OR (((forward reach*[tiab])
OR functional reach*[tiab]) OR FRT[tiab])) OR (((((((grip strength*[tiab]) OR
grip force*[tiab]) OR Hand Strength[Mesh]) OR hand strenght*[tiab]) OR Pinch
Strength[MeSH]) OR pinch strenght*[tiab]) OR hand grip*[tiab]) OR HGS[tiab])) OR

((((((((((((((((((((((abduction*[tiab]) OR abduction*[tiab]) OR adduction*[tiab]) OR adductor*[tiab]) OR adductor*[tiab]) OR ankle musc*[tiab]) OR ankle strength*[tiab]) OR calf[tiab]) OR eversion*[tiab]) OR evertor*[tiab]) OR extensor*[tiab]) OR external rotator*[tiab]) OR flexor*[tiab]) OR foot musc*[tiab]) OR foot strength*[tiab]) OR gastrocnem*[tiab]) OR glute*[tiab]) OR Hamstring Muscles[Mesh]) OR hamstring*[tiab]) OR hip musc*[tiab]) OR hip strength*[tiab]) OR internal rotator*[tiab]) OR inversion*[tiab]) OR invertor*[tiab]) OR knee musc*[tiab]) OR knee strength*[tiab]) OR leg force*[tiab]) OR leg strength*[tiab]) OR lower extremity force*[tiab]) OR lower extremity strength*[tiab]) OR Lower limb strength*[tiab]) OR perone*[tiab]) OR quadricep*[tiab]) OR soleus[tiab]) OR tibial[tiab]))) AND ((instrumentation[sh] OR methods[sh] OR Validation Studies[pt] OR Comparative Study[pt] OR «psychometrics»[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR «outcome assessment (health care)»[MeSH] OR outcome assessment[tiab] OR outcome measure*[tw] OR «observer variation»[MeSH] OR observer variation[tiab] OR «Health Status Indicators»[Mesh] OR «reproducibility of results»[MeSH] OR reproducib*[tiab] OR «discriminant analysis»[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR «internal consistency»[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab]))) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR «precise values»[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR «known group»[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR inter scale correlation*[tiab] OR error[tiab] OR errors[tiab] OR «individual variability»[tiab] OR (variability[tiab] AND

(analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR «standard error of measurement»[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR «ceiling effect»[tiab] OR «floor effect»[tiab] OR «Item response model»[tiab] OR IRT[tiab] OR Rasch[tiab] OR «Differential item functioning»[tiab] OR DIF[tiab] OR «computer adaptive testing»[tiab] OR «item bank»[tiab] OR «cross-cultural equivalence»[tiab])) NOT (((((((((((((((«addresses»[Publication Type]) OR «biography»[Publication Type]) OR «case reports»[Publication Type]) OR «comment»[Publication Type]) OR «directory»[Publication Type]) OR «editorial»[Publication Type]) OR «festschrift»[Publication Type]) OR «interview»[Publication Type]) OR «lectures»[Publication Type]) OR «legal cases»[Publication Type]) OR «legislation»[Publication Type]) OR «letter»[Publication Type]) OR «news»[Publication Type]) OR «newspaper article»[Publication Type]) OR «patient education handout»[Publication Type]) OR «popular works»[Publication Type]) OR «congresses»[Publication Type]) OR «consensus development conference»[Publication Type]) OR «consensus development conference, nih»[Publication Type]) OR «practice guideline»[Publication Type]) NOT ((«animals»[MeSH Terms] NOT «humans»[MeSH Terms]))))

REFERENCES

1. United Nations DoEaSA, Population Division (2019). United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423).
2. WHO World Health Organization report/ WHO Global report on falls prevention in older age. 2007. ISBN 978 92 4 156353 6, 2007.
3. Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC geriatrics*. 2011;11:33.
4. Iaboni A, Flint AJ. The complex interplay of depression and falls in older adults: a clinical review. *Am J Geriatr Psychiatry*. 2013;21(5):484-92.
5. Liu SW, Obermeyer Z, Chang Y, Shankar KN. Frequency of ED revisits and death among older adults after a fall. *The American journal of emergency medicine*. 2015;33(8):1012-8.
6. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E, et al. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. *Journal of geriatric physical therapy* (2001). 2017;40(1):1-36.
7. Thapa PB, Brockman KG, Gideon P, Fought RL, Ray WA. Injurious falls in nonambulatory nursing home residents: a comparative study of circumstances, incidence, and risk factors. *Journal of the American Geriatrics Society*. 1996;44(3):273-8.
8. Schoene D, Wu SM, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, et al. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. *Journal of the American Geriatrics Society*. 2013;61(2):202-8.
9. Nelson AJ, Certo LJ, Lembo LS, Lopez DA, Manfredonia EF, Vanichpong SK, et al. The functional ambulation performance of elderly fallers and non-fallers walking at their preferred velocity. *NeuroRehabilitation*. 1999;13(3):141-6.
10. Veronese N, Bolzetta F, Toffanello ED, Zambon S, De Rui M, Perissinotto E, et al. Association between Short Physical Performance Battery and falls in older people: the Progetto Veneto Anziani Study. *Rejuvenation research*. 2014;17(3):276-84.
11. Lundin H, Saaf M, Strender LE, Nyren S, Johansson SE, Salminen H. One-leg standing time and hip-fracture prediction. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2014;25(4):1305-11.
12. Chang CJ, Chang YS, Yang SW. Using single leg standing time to predict the fall risk in elderly. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference*. 2013;2013:7456-8.

13. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. *Journal of gerontology*. 1990;45(6):M192-7.
14. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *The New England journal of medicine*. 1988;319(26):1701-7.
15. 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. *Journal of gerontology*. 1994;49(2):M85-94.
16. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology)*: Cambridge: Cambridge University Press.; 2011.
17. 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. *Quality of Life Research*. 2018;27(5):1171-9.
18. 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. *Journal of clinical epidemiology*. 2010;63(7):737-45.
19. Fletcher RH, Fletcher SW, Fletcher GS. *Clinical epidemiology : the essentials*. Fifth edition: Lippincott Williams & Wilkins, a Wolters Kluwer business; 2015.
20. Pamoukdjian F, Paillaud E, Zelek L, Laurent M, Levy V, Landre T, et al. Measurement of gait speed in older adults to identify complications associated with frailty: A systematic review. *Journal of geriatric oncology*. 2015;6(6):484-96.
21. Freiburger E, de Vreede P, Schoene D, Rydwick E, Mueller V, Frandin K, et al. Performance-based physical function in older community-dwelling persons: a systematic review of instruments. *Age and ageing*. 2012;41(6):712-21.
22. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis. *BMC geriatrics*. 2014;14:14.
23. Kopke S, Meyer G. The Tinetti test: Babylon in geriatric assessment. *Zeitschrift fur Gerontologie und Geriatrie*. 2006;39(4):288-91.
24. Soares Menezes KVR, Auger C, de Souza Menezes WR, Guerra RO. Instruments to evaluate mobility capacity of older adults during hospitalization: A systematic review. *Archives of gerontology and geriatrics*. 2017;72:67-79.
25. Park SH. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. *Aging clinical and experimental research*. 2018;30(1):1-16.
26. Scott V, Votova K, Scanlan A, Close J. Multifactorial and functional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings. *Age and ageing*. 2007;36(2):130-9.
27. Aromataris E, Munn Z. (Editors). *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute, 2017
28. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ (Clinical research ed)*. 2015;351:h5527.
29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*. 2009;6(7):e1000097.
30. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. 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*. 2010;19(4):539-49.
31. Fagan TJ. Letter: Nomogram for Bayes theorem. *The New England journal of medicine*. 1975;293(5):257.
32. Hars M, Audet MC, Herrmann F, De Chasse J, Rizzoli R, Reny JL, et al. Functional Performances on Admission Predict In-hospital Falls, Injurious Falls and Fractures in Older Patients: a Prospective Study. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2018.
33. Schwendimann R, Bühler H, De Geest S, Milisen K. Characteristics of Hospital Inpatient Falls across Clinical Departments. *Gerontology*. 2008;54(6):342-8.
34. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2009;18(8):1115-23.
35. Nordin E, Lindelof N, Rosendahl E, Jensen J, Lundin-Olsson L. Prognostic validity of the Timed Up-and-Go test, a modified Get-Up-and-Go test, staff's global judgement and fall history in evaluating fall risk in residential care facilities. *Age and ageing*. 2008;37(4):442-8.
36. Muir SW, Berg K, Chesworth BM, Klar N, Speechley M. Modifiable Risk Factors Identify People Who Transition from Non-fallers to Fallers in Community-Dwelling Older Adults: A Prospective Study. *Physiotherapy Canada Physiotherapie Canada*. 2010;62(4):358-67.
37. Oliver D, Daly F, Martin FC, McMurdo ME. Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. *Age and ageing*. 2004;33(2):122-30.
38. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*. 2011;155(8):529-36.
39. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016;16(1):97.
40. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics (Oxford, England)*. 2007;8(2):239-51.

41. Macaskill P, Gatsonis CA, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0. The Cochrane Collaboration, 2010. 2010.
42. 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. *Quality of Life Research*. 2018;27(5):1147-57.
43. Greene BR, O'Donovan A, Romero-Ortuno R, Cogan L, Scanaill CN, Kenny RA. Quantitative falls risk assessment using the timed up and go test. *IEEE Trans Biomed Eng*. 2010;57(12):2918-26.
44. Marques A, Almeida S, Carvalho J, Cruz J, Oliveira A, Jácome C. Reliability, Validity, and Ability to Identify Fall Status of the Balance Evaluation Systems Test, Mini-Balance Evaluation Systems Test, and Brief-Balance Evaluation Systems Test in Older People Living in the Community. *Archives of physical medicine and rehabilitation*. 2016;97(12):2166-73.e1.
45. Muir SW, Berg K, Chesworth B, Speechley M. Use of the Berg Balance Scale for Predicting Multiple Falls in Community-Dwelling Elderly People: A Prospective Study. *Physical therapy*. 2008;88(4):449-59.
46. Santos GM, Souza AC, Virtuoso JF, Tavares GM, Mazo GZ. Predictive values at risk of falling in physically active and no active elderly with Berg Balance Scale. *Rev Bras Fisioter*. 2011;15(2):95-101.
47. Yingyongyudha A, Saengsirisuwan V, Panichaporn W, Boonsinsukh R. The Mini-Balance Evaluation Systems Test (Mini-BESTest) Demonstrates Higher Accuracy in Identifying Older Adult Participants With History of Falls Than Do the BESTest, Berg Balance Scale, or Timed Up and Go Test. *Journal of geriatric physical therapy* (2001). 2016;39(2):64-70.
48. Abu Samah Z. Discriminative and Predictive Ability of Physical Performance Measures in Identifying Fall Risk among Older Adults: Penerbit Universiti Kebangsaan Malaysia; 2018.
49. Alexandre TS, Meira DM, Rico NC, Mizuta SK. Accuracy of Timed Up and Go Test for screening risk of falls among community-dwelling elderly. *Rev Bras Fisioter*. 2012;16(5):381-8.
50. Chantanachai T, Pichaiyongwongdee S, Jalayondeja C. Fall prediction in thai elderly with timed up and go and tandem walk test: a cross-sectional study. *Journal of the Medical Association of Thailand = Chotmaihet thangphaet*. 2014;97 Suppl 7:S21-5.
51. Criter RE, Honaker JA. Identifying Balance Measures Most Likely to Identify Recent Falls. *Journal of geriatric physical therapy* (2001). 2016;39(1):30-7.
52. Criter RE, Honaker JA. Fall risk screening protocol for older hearing clinic patients. *Int J Audiol*. 2017;56(10):767-74.
53. Ibrahim A, Singh DKA, Shahar S, Omar MA. Timed up and go test combined with self-rated multifactorial questionnaire on falls risk and sociodemographic factors predicts falls among community-dwelling older adults better than the timed up and go test on its own. *J Multidiscip Healthc*. 2017;10:409-16.
54. Kojima G, Masud T, Kendrick D, Morris R, Gawler S, Trembl J, et al. Does the timed up and go test predict future falls among British community-dwelling older people? Prospective cohort study nested within a randomised controlled trial. *BMC geriatrics*. 2015;15:38.
55. Shimada H, Suzukawa M, Tiedemann A, Kobayashi K, Yoshida H, Suzuki T. Which neuromuscular or cognitive test is the optimal screening tool to predict falls in frail community-dwelling older people? *Gerontology*. 2009;55(5):532-8.
56. Trueblood PR, Hodson-Chennault N, McCubbin A, Youngclarke D. Performance and Impairment-Based Assessments Among Community Dwelling Elderly: Sensitivity and Specificity. *Journal of Geriatric Physical Therapy*. 2001;24(1):2-6.
57. Arnold CM, Faulkner RA. The history of falls and the association of the timed up and go test to falls and near-falls in older adults with hip osteoarthritis. *BMC geriatrics*. 2007;7:17.
58. Rose DJ, Jones CJ, Lucchese N. Predicting the Probability of Falls in Community-Residing Older Adults Using the 8-Foot Up-and-Go: A New Measure of Functional Mobility. *Journal of Aging and Physical Activity*. 2002;10(4):466.
59. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Physical therapy*. 2000;80(9):896-903.
60. Murphy MA, Olson SL, Protas EJ, Overby AR. Screening for falls in community-dwelling elderly. *Journal of Aging and Physical Activity*. 2003;11(1):66-80.
61. Middleton A, Fulk GD, Herter TM, Beets MW, Donley J, Fritz SL. Self-Selected and Maximal Walking Speeds Provide Greater Insight Into Fall Status Than Walking Speed Reserve Among Community-Dwelling Older Adults. *Am J Phys Med Rehabil*. 2016;95(7):475-82.
62. Viveiro LAP, Gomes GCV, Bacha JMR, Carvas Junior N, Kallas ME, Reis M, et al. Reliability, Validity, and Ability to Identify Fall Status of the Berg Balance Scale, Balance Evaluation Systems Test (BESTest), Mini-BESTest, and Brief-BESTest in Older Adults Who Live in Nursing Homes. *Journal of geriatric physical therapy* (2001). 2019;42(4):E45-e54.
63. Applebaum EV, Breton D, Feng ZW, Ta AT, Walsh K, Chassé K, et al. Modified 30-second Sit to Stand test predicts falls in a cohort of institutionalized older veterans. *PloS one*. 2017;12(5):e0176946.
64. Chow RB, Lee A, Kane BG, Jacoby JL, Barraco RD, Dusza SW, et al. Effectiveness of the "Timed Up and Go" (TUG) and the Chair test as screening tools for geriatric fall risk assessment in the ED. *The American journal of emergency medicine*. 2019;37(3):457-60.
65. Gafner SC, Bastiaenen CH, Ferrari S, Gold G, Terrier P, Hilfiker R, et al. Hip muscle and hand-grip strength to differentiate between older fallers and non-fallers: a cross-sectional validity study. *Clinical interventions in aging*. 2018;13:1-8.
66. Sousa LM, Marques-Vieira CM, Caldevilla MN, Henriques CM, Severino SS, Caldeira SM. Risk for falls among community-dwelling older people: systematic literature review. *Revista gaucha de enfermagem*. 2017;37(4):e55030.

67. Ganz DA, Higashi T, Rubenstein LZ. Monitoring Falls in Cohort Studies of Community-Dwelling Older People: Effect of the Recall Interval. *Journal of the American Geriatrics Society*. 2005;53(12):2190-4.
68. Cummings SR, Nevitt MC, Kidd S. Forgetting falls. The limited accuracy of recall of falls in the elderly. *Journal of the American Geriatrics Society*. 1988;36(7):613-6.
69. Hoffman GJ, Ha J, Alexander NB, Langa KM, Tinetti M, Min LC. Underreporting of Fall Injuries of Older Adults: Implications for Wellness Visit Fall Risk Screening. *Journal of the American Geriatrics Society*. 2018;66(6):1195-200.
70. Faber MJ, Bosscher RJ, van Wieringen PC. Clinimetric properties of the performance-oriented mobility assessment. *Physical therapy*. 2006;86(7):944-54.
71. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine*. 2000;45(1-2):23-41.
72. Stevens JA, Phelan EA. Development of STEADI: A Fall Prevention Resource for Health Care Providers. *Health Promotion Practice*. 2013;14(5):706-14.
73. Wang-Hsu E, Smith SS. Interrater and Test-Retest Reliability and Minimal Detectable Change of the Balance Evaluation Systems Test (BESTest) and Subsystems With Community-Dwelling Older Adults. *Journal of geriatric physical therapy* (2001). 2018;41(3):173-9.
74. Anson E, Thompson E, Ma L, Jeka J. Reliability and Fall Risk Detection for the BESTest and Mini-BESTest in Older Adults. *Journal of geriatric physical therapy* (2001). 2019;42(2):81-5.
75. Hofheinz M, Schusterschitz C. Dual task interference in estimating the risk of falls and measuring change: a comparative, psychometric study of four measurements. *Clin Rehabil*. 2010;24(9):831-42.
76. Lee S-P, Dufek J, Hickman R, Schuerman S. Influence of Procedural Factors on the Reliability and Performance of the Timed Up-and-go Test in Older Adults. *International Journal of Gerontology*. 2016;10(1):37-42.
77. Smith E, Walsh L, Doyle J, Greene B, Blake C. The reliability of the quantitative timed up and go test (QTUG) measured over five consecutive days under single and dual-task conditions in community dwelling older adults. *Gait & posture*. 2016;43:239-44.
78. Hartmann A, Murer K, de Bie RA, de Bruin ED. Reproducibility of spatio-temporal gait parameters under different conditions in older adults using a trunk tri-axial accelerometer system. *Gait & posture*. 2009;30(3):351-5.
79. Goldberg A, Schepens S. Measurement error and minimum detectable change in 4-meter gait speed in older adults. *Aging clinical and experimental research*. 2011;23(5-6):406-12.
80. Hars M, Herrmann FR, Trombetti A. Reliability and minimal detectable change of gait variables in community-dwelling and hospitalized older fallers. *Gait & posture*. 2013;38(4):1010-4.
81. Arnold CM, Warkentin KD, Chilibeck PD, Magnus CR. The reliability and validity of handheld dynamometry for the measurement of lower-extremity muscle strength in older adults. *J Strength Cond Res*. 2010;24(3):815-24.
82. Ford-Smith CD, Wyman JF, Elswick RK, Jr., Fernandez T. Reliability of stationary dynamometer muscle strength testing in community-dwelling older adults. *Archives of physical medicine and rehabilitation*. 2001;82(8):1128-32.
83. Katoh M, Isozaki K. Reliability of Isometric Knee Extension Muscle Strength Measurements of Healthy Elderly Subjects Made with a Hand-held Dynamometer and a Belt. *Journal of physical therapy science*. 2014;26(12):1855-9.
84. André HI, Carnide F, Borja E, Ramalho F, Santos-Rocha R, Veloso AP. Calf-raise senior: a new test for assessment of plantar flexor muscle strength in older adults: protocol, validity, and reliability. *Clinical interventions in aging*. 2016;11:1661-74.
85. Hofheinz M, Mibs M. The Prognostic Validity of the Timed Up and Go Test With a Dual Task for Predicting the Risk of Falls in the Elderly. *Gerontol Geriatr Med*. 2016;2:2333721416637798.
86. Holbein-Jenny MA, Billek-Sawhney B, Beckman E, Smith T. Balance in personal care home residents: a comparison of the Berg Balance Scale, the Multi-Directional Reach Test, and the Activities-Specific Balance Confidence Scale. *Journal of geriatric physical therapy* (2001). 2005;28(2):48-53.
87. Martinez BP, Batista AK, Ramos IR, Dantas JC, Gomes IB, Forgiarini LAJ, et al. Viability of gait speed test in hospitalized elderly patients. *J Bras Pneumol*. 2016;42(3):196-202.
88. Bautmans I, Jansen B, Van Keymolen B, Mets T. Reliability and clinical correlates of 3D-accelerometry based gait analysis outcomes according to age and fall-risk. *Gait & posture*. 2011;33(3):366-72.
89. Bruyneel AV, Gafner SC, Ferrari S, Gold G, Monnin D, Terrier P, et al. Intra-rater reliability of hip abductor isometric strength testing in a standing position in older fallers and non-fallers. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity*. 2018;15:9.
90. Gafner S, Bastiaenen CHG, Terrier P, Punt I, Ferrari S, Gold G, et al. Evaluation of hip abductor and adductor strength in the elderly: a reliability study. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity*. 2017;14:5.

CHAPTER 8

General Discussion

“Aging in place” is a wish many of us have while getting older (1). Being able to take care of one’s own garden and pets despite getting older is a privilege. Independently making one’s own tea, coffee, or meals, taking pleasure in reading a book and the newspaper or doing some handicraft in one’s favorite comfortable chair are often linked to a good quality of life, and they are among the things we wish for our loved ones.

Persons aged ≥ 65 years are currently the world’s fastest-growing age group (2). With increasing age comes an increasing risk of having a first fall or recurrent falls as well as associated injuries and fractures. Falls can be a source of loss of independence (3), hospitalizations (4), nursing home admissions (5) and increased mortality (4), and fall risk detection has thus become a global public health challenge.

As already clear by the aim of this thesis, it is important that clinicians can detect older persons at risk of falls as early as possible to initiate the necessary measures to prevent a fall and/or postpone recurrent falls.

Multiple intrinsic and extrinsic factors influence the risk of falling (e.g., history of falls, polypharmacy, incontinence, change of cognitive function, fear of falling and low education) (6). Among these, modifiable risk factors that are open to interventions in prevention and rehabilitation are of great interest. Consequently, numerous attempts have been made to develop functional tests to identify older persons at high risk of falls and fractures. Most of these functional tests do not provide specific information on the underlying cause influencing the risk of falls that should be targeted in prevention or rehabilitation. In addition, we observe that information about clinimetric properties of commonly used functional tests are not easily accessible for clinicians and the knowledge of their interpretation is only limited. Therefore, one of the projects integrated in this thesis is a systematic review evaluating intra-rater and inter-rater reliability and diagnostic accuracy of currently used fall-risk assessments. Our goal was to present the results for clinicians, facilitating their decision-making concerning the optimal choice for a suitable fall-risk assessment for older persons, and to explain the use of LRs and post-test probabilities that should influence this choice. Furthermore, information on the diagnostic accuracy of currently used functional fall-risk assessments is essential in relation to the subsequent main part of this thesis. It was important to compare the developed fall risk parameter with currently used fall-risk assessments and to investigate their potential to be used in combination to increase their diagnostic accuracy.

The thesis is focused on the evaluation of hip muscle strength, more specifically hip abductor strength. Hip abductor strength (and hip muscle strength in general) is an easily targetable parameter in prevention and rehabilitation. We started this PhD trajectory by assessing the value of hip frontal plane strength in relation to fall risk (7-9). Hip frontal plane strength is, based on previous studies, important to

stabilize mediolateral balance in older persons (10-12), and it can compensate for decreased ankle proprioceptive thresholds such as in persons with polyneuropathy due to diabetes (13). Allet et al. (13) concluded that the compensation of hip abductor strength for decreased proprioceptive thresholds is a clinically relevant finding because hip muscles can be strengthened even in older patients with polyneuropathy and might as such improve balance (unipedal stance time) (13). Based on these facts and knowing that hip abductor muscle strength shows a more pronounced decline in older persons compared to hip extensor strength (14), focusing on hip abductor strength is of high clinical relevance.

Several aspects of the hip abductor strength assessment have been investigated to be able to recommend or reject its evaluation in relation to the detection of fall risk of older persons. The following paragraphs summarize those important aspects investigated during this PhD thesis.

MAIN FINDINGS

Reliability and feasibility of hip abductor strength testing

In a first step, we assessed the relative and absolute reliability, a prerequisite for good validity, and the feasibility of a hip abductor and adductor strength test in a side-lying and standing position (Chapters 2 and 3). A standardized way to measure hip abductor and adductor strength was used in the first study (8). The participant was in a side-lying position and pushed his or her leg as hard and fast as possible against a dynamometer fixed to a custom-made, metallic frame that was fixed to the treatment table (8). The choice of the side-lying position was influenced by Widler et al. (15). They found in young participants (mean age 31 ± 6) that the hip abduction strength assessment in side-lying position showed the highest maximum strength and the least compensation with the gluteus medius of the contralateral side compared to supine and standing position (15). Our results showed that hip abductor and adductor MVIS and RFG assessment in side-lying position is feasible and shows high values for reliability in persons 65 and older (80.46 ± 7.05) (8). However, the assessment time in side-lying position is too long for daily clinical use (8). This led to the need to reconsider the test position and to assess the feasibility and reliability of the hip abductor strength assessment in a standing position. To exclude the influence of the tester’s strength on the measurements, a dynamometer was manually fixed against a wall at a height at which the participant could push with their malleolus externus against the dynamometer (9). It seems logical that most falls occur when individuals are in a standing position or moving around. Therefore, a standing position for hip abductor strength assessment is much more functional

than a side-lying assessment of hip abductor strength is. Our study results showed that hip abductor strength assessment in a standing position is feasible and achieves $ICC_{\text{agreement}}$ values that are as high as the ones assessed in a side-lying position (Chapter 3) (9).

Because reliability parameters dependent highly on the heterogeneity of the study sample (16), we chose to present our results in the total study group as well as in more homogenous sub-groups of fallers and non-fallers separately and found high reliability for the sub-groups. The test of hip abductor strength in a standing position takes about half the time the hip abductor strength assessment takes in a side-lying position and hence seems much more feasible in clinical practice (8, 9). For clinical use, this is encouraging because it seems that hip abductor strength can be assessed in a functional standing position without additional fixation of the dynamometer and with easily accessible material in everyday clinical practices.

Although both testing positions showed a good intra-rater reliability, in both cases, high standard error of measurement (SEM) and smallest detectable difference (SDD) were evaluated, with values of SDD that are slightly lower in the standing than in the side-lying position (8, 9). According to the COSMIN group, the SDD is defined as the change beyond the measurement error; this is a change that falls outside the limits of agreement of the Bland and Altman method (17). Therefore, the SDDs shown in our reliability results (8, 9) show how large a difference of two measurements must be to detect a real difference, outside measurement error. In our articles, these values are high, reporting high values of measurement error.

Assessing the reliability and feasibility of a measurement procedure is important and an essential requirement when developing a new measurement tool (17). For clinical use, it is important to know what the range of scores for patients who have not changed will be for different raters at the same time (inter-rater reliability) or by the same rater on different occasions (intra-rater reliability) (18). Therefore, good reliability is a prerequisite for research regarding validity. Measurement instruments can be used for discriminative, evaluative and prospective purposes, and their validity must be assessed accordingly (17). Regarding the hip abductor strength test, the ability to discriminate between fallers and non-fallers who are 65 and older is of importance. Therefore, in the next steps after having investigated the intra-rater reliability, we evaluated the discriminative ability of first all hip muscles strength groups, followed by a focus on the diagnostic accuracy of hip abductor strength.

Criterion validity: Hip muscle strength, hip abductor strength and the development of a new measurement tool and measuring procedure

To be able to answer the question about the importance of hip abductor strength in relation to fall risk, the first step was to evaluate which hip muscle group best discriminates between older fallers and non-fallers (Chapter 4) (7). It turned out that hip abductor strength was the only assessed hip muscle strength group that showed a discriminative ability that was predefined as good enough for clinical use ($AUC > 0.7$), as compared to the chosen “gold standard” fall history, and it corresponds to the criteria for good measurement properties proposed by the COSMIN group (19, 20). The discriminative ability of hip abductor strength was also higher than that of hand-grip strength, which we included in the analyses because it is a parameter that is assumed to be related to whole body strength and fall risk (21). This study underlined the importance of hip abductor strength in relation to older persons’ fall risk (7). Hence, we continued with the evaluation of the specific psychometric properties of the hip abductor strength (Chapter 5). We assessed their diagnostic accuracy, compared it to the TUG and SPPB of the same population, and found that they are equivalent. Because falls are influenced by many different factors, we were further interested in net sensitivities and specificities of a combined assessment of hip abductor strength together with TUG and SPPB and found high net sensitivities confirming our hypotheses (22). One might think this is rather surprising because the hip abductor strength test only assesses one specific parameter out of the fall-influencing factors. However, the results of the criterion validity (discriminative ability between fallers and non-fallers $AUC > 0.7$) (7), as well as the encouraging diagnostic accuracy of the hip abductor strength test comparable to currently used fall risk tests (22), which also fulfills the prerequisite of high intra-rater reliability (8, 9), is an indication of the importance of hip abductor strength in relation to falls (22). These findings are encouraging in our reasoning behind hip abductor strength as a component within a test battery of fall risk in daily clinical use.

However, we were confronted with more challenges in relation to the assessment position as well as the measurement instrument and its functional and mostly clinical applicability throughout the mentioned studies. Those challenges, as well as feedback from and discussions with health care professionals in clinics and at conferences, led us to reflect and inspired a way to assess hip abductor strength in daily clinical life. Therefore, with the focus on allowing our previous findings to be integrated in clinics, the additional project for the development of a measurement procedure in standing position emerged for this thesis.

Falls mostly occur while older persons are moving around (e.g., while standing up, walking or turning around) (23-25). Hip abductors are known to stabilize the

upper body over the support leg and counteract destabilization in the mediolateral direction (26). Therefore, they prevent the center of mass from rapidly falling downward and laterally towards the swing side (26). Consequently, a test position in a standing, closed chain position, measuring the hip abductor strength of the support leg seems much more functional and directly related to a potential fall risk than the measurement positions laying on the side or standing, measuring the hip abductor strength of the swing leg. If hip abductor strength is a fall-influencing parameter, then falls probably occur related to weak hip abductors of the weight bearing leg rather than of the lifted one. With the assistance of engineers of the Institute of Industrial Technologies and Sciences, University of Applied Sciences and Arts Western Switzerland (HES-SO Geneva/HEPIA), this reasoning led to the development of a prototype of an easy to manipulate and transportable measurement instrument and associated measurement procedure (patent was filed at the European patent office; application and patent number 202000 13.9-1115 [impact paragraph]) for assessing hip abductor strength in a standing, closed-chain position. With the goal of evaluating the adapted measurement procedure to allow a later use in daily clinical life, the study protocol of a reliability and validity study with the adapted measurement instrument and procedure was developed (Chapter 6), the ethical approval for the study was obtained and the recruitment of the first participants started. Methodological considerations of our previous projects were integrated in the study protocol (e.g., blinding of assessors, assessment of inter- and intra-rater reliability, and prospective fall assessment) (Chapter 6) (27).

Reliability and diagnostic accuracy of commonly used performance tests relative to fall history in older persons

Conscious of the multifactorial etiology of falls (6) and aware that assessing hip abductor strength as a fall risk parameter assesses only one specific factor of the “body functions and structures” component according to the international classification of functioning (ICF) (28), we were interested in the diagnostic properties of other fall-risk assessments. We assumed that we have to combine the assessment of hip abductor strength with other performance-based measurements for a multifactorial fall-risk assessment. Consequently, we assessed in a systematic review the clinimetric properties of other tests that assess body functions and structures and/or the ICF “activities and participation” components (28) in relation to fall risk of older persons. The systematic review aimed to provide an overview of the reliability (inter- and intra-rater) and diagnostic accuracy, specifically post-test probability, of commonly used performance tests to assess the fall risk of persons older than 65 according to their setting (community dwelling, nursing home, or hospital; Chapter

7). The overview can support caregivers in making decisions with the person at risk of falls to help them choose a suitable fall-risk assessment corresponding to the patient’s specific setting.

Over all settings, the (intra- and inter-rater) reliability of the different measures was, except for one test, high enough to be recommended for clinical use. For the diagnostic accuracy, different results were found for each setting, and a recommendation based on the outcomes regarding reliability and diagnostic accuracy was made (Chapter 7). However, due to different limitations, these recommendations are not very strong.

Different aspects made this systematic review a challenging and time-consuming project. Our purposely extensive literature search led to a high number of identified articles. The presentation of post-test probabilities induced the creation of 2 x 2 tables, including the numbers of the true positive and true negative persons and allowing the calculations or presentations of sensitivities, specificities and LR_s. Many diagnostic accuracy studies do not present the necessary data very clearly to create the 2 x 2 tables (e.g., missing detailed description of the index and/or reference test to be able to judge how persons were classified, numbers of true positives and true negatives not presented, and total number of tested persons or sub-groups not stated). Hence, we recommend researchers follow specific guidelines such as the Standards for Reporting of Diagnostic Accuracy Studies (29) while planning and reporting diagnostic accuracy studies. This will facilitate researchers and clinicians’ use of those studies and the clinical application because, for example, post-test probabilities can be evaluated and help influence the right choice for a measurement instrument. For clinical use, we also encourage researchers always to present the cut-off values for which the diagnostic accuracy was calculated. Additional challenges faced during this systematic review (evaluation method of cut-off values, chosen reference standard) are further described in the methodological consideration part of this thesis.

Reflections on strengths and limitations

This dissertation has strengths and limitations. This PhD project is the first that investigates the clinimetric properties of hip abductor strength compared to the other hip muscle groups and their ability to discriminate between fallers and non-fallers. Many facts indicate the relevance of hip abductor strength in relation to fall risk. Therefore, this dissertation investigates hip abductor strength as a promising contributing factor for assessing fall risk alone or in a test battery with a strong focus on the clinical applicability of the results. The information about the good intra-rater reliability, the established cut-off values and the associated diagnostic accuracy

allows direct clinical use of our findings. The development of a new measurement procedure (Chapter 6) that emerged from the Chapters 2 to 5 is a further strength of this thesis on the journey to an easy and transportable way of a functional assessment of hip abductor strength and their clinical implementation for fall-risk assessment. The specific information about this variable also allows its well-targeted treatment after additional investigations in future research.

The additional examination of the clinimetric properties of currently used fall-risk assessment in our systematic review is a further strength. The results of Chapter 7 give clinicians an overview of the intra- and inter-tester reliability and the diagnostic accuracy, with a focus on the post-test probability, and help facilitate their choice of a suitable fall-risk assessment for older persons in their specific setting. The explanations of how to use and interpret the different values of the diagnostic accuracy reinforces clinicians' improved understanding of those values and their ability to use their interpretation of the specific test in daily clinical life, with the patient as the central figure who will benefit. Additionally, the systematic review shows that currently used fall-risk assessments do not have a diagnostic accuracy that allows their sole use to obtain information about the fall risk of older persons but that a combination of more than one test should be encouraged. The limitations are discussed in Chapters 2 to 7 and in the following methodological reflections. Nevertheless, we would like to point out that our rather small sample including in- and outpatients, although it is based on a sample size calculation, might have led to an over- or underestimation of the results of the diagnostic accuracy. Diagnostic accuracy studies are strongly influenced by the studied sample and the interpretation of the results, and their implementation into clinical practice always must be done while keeping this fact in mind (e.g., similar patient characteristics, setting). The study protocol for the adapted measure (Chapter 6) takes into account the limitations of our previous projects and thus includes a much larger sample size and a prospective study design.

Reflections on methodological aspects

As in all research projects, some methodological choices of this dissertation can be discussed and merit a deeper reflection.

Evaluation of cut-off values and associated diagnostic accuracy of a test

Approaching the end of this dissertation, we now certainly know that falls can cause potentially serious consequences, which is the reason our studies we sought a

test with a high ability to detect true positives (high sensitivity) to be as certain as possible of identifying persons at risk of falls (true positives). Therefore, the diagnostic accuracy of hip abductor strength in our study was evaluated at a 90% sensitivity level (22). It was judged more important to be sure to detect as many persons as possible at risk of falls, taking the "risk" to have a lower specificity and therefore more false positives. Clinically, it seemed to be less serious to include older persons who are not at risk of falls into a fall prevention program than not to include someone who is at risk. Moreover, because we know that the risk of falls is high once persons reached the age of 65 (annual fall rate of 28–35% for community-dwelling persons, 30–50% for people living in long-term care institutions (30)). Other authors chose another methodology to evaluate a cut-off value and calculated the Youden Index, which is the highest possible combination between sensitivity and specificity a test reaches (point on the upper left of the ROC) (31). The Youden Index gives equal weight to sensitivity and specificity and facilitates the comparison of different diagnostic tests from that perspective (31). Nonetheless, this implies that the "costs" of false-positive and false-negative results are equivalent (31), which we did not judge to be the case in our study and which from a clinical perspective is not often the case. Other studies have chosen to investigate the diagnostic accuracy of an assessment at predefined cut-off values that were used in earlier studies (32). The choice of the 90% sensitivity or other methods to evaluate the cut-off values influences the diagnostic accuracy of measurement instruments. This is an important point to keep in mind when different studies and assessments are compared.

Falls have multifactorial causes and we know that no fall-risk assessment in a single use has high diagnostic accuracy (33). Additionally, currently used fall-risk assessments usually do not indicate which specific parameter should be targeted in an intervention. From a clinical perspective, this indicates that different fall-risk assessment tools must be combined to obtain better insight into the fall risk of our patients and to detect the specific components that need be targeted for an effective prevention and/or rehabilitation program. The combined use of different fall-risk assessments can be performed in parallel testing (also called "simultaneous testing", leading to an increase of the sensitivity, and net sensitivity) or serial testing (also called "sequential testing", increase of the specificity, and net specificity) (34). As in the case of fall-risk assessment, the precise detection of persons at risk (true positives) is important, we want to use assessments that achieve high sensitivity. Therefore, when tests have to be combined for fall risk detection, parallel testing should be favored. Parallel testing with different tests means that several tests are performed together at the same time, and when one or more of them are positive, the test result is judged as positive (person is at risk of falls) (34).

In our study, the net sensitivity of hip abductor strength in parallel testing with the SPPB and TUG was calculated and showed very high values, which is encouraging for our hypothesis of hip abductor strength as an important parameter in relation to fall risk of older persons. However, the presentation of net sensitivities or net specificities is not very common in current literature but would be helpful for clinicians in their choice of different measurement instruments. The choice to use either parallel or serial testing should be based on different considerations of the objectives of the tests such as whether they should be used for screening or diagnostics and practical reasons such as the setting in which the tests will be performed (34).

Fall history as the reference test

The choice of fall history as the reference standard merits a closer look. Previous falls are a known risk factor for future falls (35). A fall is often defined as an event resulting in a person inadvertently coming to rest on the ground, floor, or other lower level (30). Many studies use fall history as the reference standard (33, 36, 37). However, some risk of recall bias exists when participants are asked if they fell in the previous months (38) and it is known that falls are likely to be underreported by older adults (39, 40). Evidently, this contains a risk of misclassification of participants into the fallers or non-fallers group, influencing the diagnostic accuracy of a test. Due to the recall bias, falls should ideally be assessed in a prospective way. However, in our systematic review (Chapter 7 (41)) and other reviews (33, 36), much more studies were found with assessing fall history as a reference standard compared to the number of falls assessed in a prospective manner. Again, this could lead to an over- or underestimation of the diagnostic accuracy of different fall-risk assessments. Additionally, different classifications of fallers versus non-fallers are used. Some studies count persons with zero to one fall as non-fallers and ≥ 2 falls as fallers (42) and others classify only people with no falls in the non-faller group and persons with ≥ 1 falls as fallers (43, 44). In our studies, the latter definition was used for the inclusion and exclusion of the articles. Other studies specifically exclude hazardous falls (e.g., when a chair collapses) (45) others do not mention whether they were included or excluded. Studies comparing non-fallers to recurrent or even injurious fallers exist as well (46). All those differences clearly can influence the diagnostic accuracy due to differences in the classifications of the fallers and non-fallers, and clinicians should be careful to check those facts when interpreting results.

The use of fall history as a reference standard also implies that the definition of a fall is well explained to participants. If participants have other personal definitions

of a fall, then this could be a further source of misclassification and, therefore, have an effect on the diagnostic accuracy of the investigated measurement tool.

Our project was the first to investigate the ability of the strength of all hip muscle groups, and later specifically the hip abductor strength, in relation to their discriminative ability for older fallers and non-fallers. In this first step of their evaluation, we used the most often-used reference standard, falls or no falls, to investigate their diagnostic accuracy. Due to the recall bias and the different definitions of a fall the patients use in clinics to avoid declaring that they fell, we opted for categories that are clearly different from each other. It only was during the progress of all the included projects and mostly during the data extraction of the systematic review that we “stumbled” across “near-falls”. From a clinical perspective, it is important to start treating persons at risk of falls as soon as possible in the hope of being able to prevent a first fall. Specific questioning about near falls thus could be interesting as an indication of their nascent insecurities and possibly their increased risk of falls.

In the context of our projects, fall history was the most appropriate reference standard, but it seems important to keep in mind the above-mentioned challenges that it brings for the evaluation of diagnostic accuracy.

Combined assessment of hip abductor maximum voluntary isometric strength and rate of force generation

A further important methodological point to discuss is the evaluation of the hip abductor maximum voluntary isometric strength (MVIS) and rate of force generation (RFG) in the same trial. Regarding the diagnostic accuracy values, hip abductor MVIS always shows slightly higher values than abductor RFG. For a layman, it might be expected that RFG would be more important for the prevention of a fall than MVIS because, for example, a fast recovery step is needed to regain balance after tripping (47). To be able to produce a rapid rise of force during the first 300 ms of a contraction is as important for older individuals counteracting sudden perturbations in balance as it is for trained athletes (48). Nevertheless, it is known that movement and information processing slow down with advancing age, as more importance is probably given to movement accuracy than movement speed (49). This might lead to speculations that MVIS is more important in relation to fall risk than RFG is. One of the early articles that inspired our work found a difference for the RFG of the hip ABD between older fallers and non-fallers (50). However, because those changes occurred in the late phase of the contraction (at 150 ms and 200 ms), they associate this difference between fallers and non-fallers more to the MVIS than the RFG, which would correspond to our finding of the higher diagnostic accuracy of hip abductor MVIS than the RFG. The fact that RFG later than 90 ms after the onset of contraction

is more related to MVIS is also stated by Andersen et al. (51). Based on their article, we calculated the RFG over 50 ms after 10% of the MVIS was reached because we intended to measure two different components of strength (MVIS and RFG).

However, the small differences between the diagnostic accuracy of abductor MVIS and abductor RFG make it difficult to know whether MVIS is clearly better for discriminating between fallers and non-fallers than RFG is or if the former is influenced, for example, by how we assessed RFG. We reconsidered assessing both MVIS and RFG in the same trial, given that telling the persons to push as hard and as fast as possible against the dynamometer could be too difficult for our participants, with an average age of 82 years. We recommend separating the assessments of MVIS and RFG in future studies to evaluate if MVIS or RFG has better or equal discriminating abilities for older fallers and non-fallers.

Another challenge for assessing the RFG was that when the subsequent projects were developed, no dynamometer was available with a solution high enough to register the RFG. Therefore, we coupled an analog dynamometer able to measure forces between 0 to 667 N (SENSIX®, Poitiers, France), with a precision of 0.002 N, with the DELSYS System (Trigno sensor, DELSYS, Inc., Boston, MA), which digitalized the analog output (3.3 V) with a sampling rate of 1926 Hz and a 16-bit resolution. The system worked well and corresponded to our expectations. However, it still could have been the source of some errors, mostly in assessing the RFG, due to the sensibility of the RFG measures to noise in the signal (48).

Due to all these aspects, we would now choose to separate the trials for MVIS and RFG and evaluate their diagnostic accuracy in separate strength assessments.

Implications for clinical practice

Evaluating and understanding the construct of older persons' fall risks benefit from the additional assessment of hip abductor strength, as a parameter that is not yet systematically assessed in current fall-risk assessments evaluating complementary aspects of falls (22). Health care professionals can assess hip abductor strength easily in a functional position, to add value to the currently used fall-risk assessments and in fall-risk prevention and rehabilitation programs.

Assessing muscle strength and power as well as its preservation with advancing age has great clinical importance because good muscle strength and power diminish the risks of falls, fractures and associated adverse events (52) and therefore contribute to maintaining independence. Nevertheless, as Stevens et al. (53) state, assessing fall risk alone does not decrease falls (53). An effective individualized intervention provided by health care professionals that includes supplemental educational materials must follow the fall-risk assessment (see implications for future

research) (53). Our systematic review of currently used fall-risk assessments provides an overview of their reliability (inter-tester and test-retest) and diagnostic accuracy. The results can help clinicians to choose the correct measurement instrument for their specific purpose and setting. The chosen fall-risk assessments generally did not show very good post-test probabilities, or the assessments showing good post-test probabilities were supported mostly by single studies with low sample sizes. Therefore, we cannot make a firm recommendation of one single fall-risk assessment alone to evaluate the fall risk of a person aged ≥ 65 . This indicates that several tests must be used in combination for clinical use. This increases the amount of time needed to classify persons as fallers or non-fallers, which can be challenging for clinicians because they have only a limited amount of time available for treatments. Therefore, for clinicians to invest their scarce treatment time well, it is even more important that overviews of clinometric properties of fall-risk assessments and their respective combinations be available to facilitate choosing the most suitable test combinations.

Another challenging but expected finding of our systematic review is that many cut-off values are used for the same measurement instrument within the same setting, rendering the clinical use of fall-risk assessments complex. Therefore, clinicians must understand which parameters are influencing diagnostic accuracy and, thus, how to interpret the results of specific articles. To receive diagnostic accuracy similar to what was published, it is important to apply the specific tests and cut-off values in a population that resembles the study population as much as possible. Summary receiver operating characteristic (SROC) curves (54) can be calculated and optimal cut-off values can be evaluated as indications of a test's global performance. These optimal cut-off values might help clinicians to choose the cut-off value for use in their clinical setting.

To facilitate making the information available for health care professionals, we present post-test probabilities in our systematic review and explain their use and how they can be calculated. However, with the very wide range of a) existing measurement instruments, b) existing settings in which they are used and c) cut-off values, choosing tests based on their clinometric properties and also interpreting the test results still are confusing for health care professionals. In our systematic review (Chapter 7), we explain important aspects of a measurement instrument, their influencing factors and their interpretation for clinicians.

Due to the vast number of existing cut-off values, which forced us to calculate the theoretical optimal cut-off value, we recommend that clinicians use the test results as a probability of fall risk, rather than a dichotomised yes or no value. The decision to categorize an older person as a faller or non-faller should be a process between the health care professional and the patient considering the results of several fall-

risk assessments, the patient's setting, their personal goals and other information. For example, the Stopping Elderly Accidents, Deaths, and Injuries (STEAR) fall-prevention toolkit suggests it might be a good idea to ask the participant some basic questions (about their fall history, fear of falling and unsteadiness while standing or walking), and perform three basic fall-risk assessments (e.g., TUG, 30-second chair stand and four-step balance test) for a general assessment of fall risk (53). Thereafter, different fall-influencing factors should be assessed to obtain targeted information about which fall risk factors to treat (53).

The currently used multidimensional fall-risk tests are easily practicable in different settings and are commonly used to assess fall-risk assessment. We recognize the strength of these tools. However, a low score on these tests (weak performance) gives no strong indication for the underlying cause for falling—for example, it does not indicate which specific body functions should be taken care of, trained and/or treated. Indeed, a weak performance on these scales could be due to a variety of cognitive, neuromuscular and/or musculoskeletal variables. Consequently, only non-specific interventions often are recommended, without explicit recommendations on which (and how) specific neuromuscular or neuro-cognitive functions should be trained. We recommend that clinicians use these tests as the first step of assessing fall risk, as described in the STEADI fall prevention toolkit (53). The single performance tests (e.g., gait speed, single leg stance, functional reach test, grip strength, lower limb muscle strength) we investigated in our systematic review as well as others might be classified within the specific tests that are recommended to investigate the specific underlying causes of fall risk. Hip abductor strength should also be classified within these specific tests.

However, fall risk is influenced by many factors, and a proper fall-risk assessment must include different fall-risk influences. However, our studies confirm the importance of evaluating hip abductor strength, which is not yet systematically assessed, in addition to currently used fall-risk assessments evaluating the complementary aspects of falls (7-9, 22).

Assessing diagnostic accuracy has many challenging aspects, and it is not always easy to provide an overview of all possible options and variables to evaluate in research projects. However, the most important point is to keep one's orientation as a researcher in the labyrinth of diagnostic accuracy and to transfer this knowledge in a way that it can be used directly in clinical practice. Researchers must make their work understandable for direct application to clinical life. It is very important to close the gap between research and practice in order to progress together and develop best-practice guidelines for patients.

Recommendations for future research

In our studies examining hip abductor strength, we used fall history over the previous 12 months as the reference standard and obtained encouraging results for including hip abductor strength in a test battery for fall-risk assessment of persons aged ≥ 65 . However, a prospective study should examine the discriminative ability of hip abductor strength in which falls are monitored closely over a follow-up time of, for example, 12 months. Additionally, most of the articles included in our systematic review used fall history, but only a few prospective articles could be included. To obtain stronger results with which to facilitate clinicians' choice of measurement instruments, we recommend performing diagnostic studies in a prospective manner with a well-chosen methodology and correct sample sizes.

Our systematic review (Chapter 7) showed that most of the included studies were performed in a community-dwelling setting. However, it seems important for future research to evaluate fall-risk assessments for a frailer population in nursing homes and in a hospital setting. The currently used fall-risk assessments might be too challenging for these populations, and other more adapted tests should be used and implemented. Furthermore, we recommend that future researchers do not combine participants from different settings due to influences from the prevalence of a disease/disorder on the diagnostic accuracy as well as potential spectrum bias.

To be effective an individualized intervention provided by health care professionals, together with supplemented educational material must follow the fall-risk assessment (53). In the different projects of this PhD thesis, we investigated, for the first time, which hip muscle strength group can best differentiate between older fallers and non-fallers as well as the specific reliability and diagnostic accuracy of measures of hip abductor strength (MVIS and RFG). Knowledge of the importance and capability of hip abductor strength to discriminate between fallers and non-fallers is an important first step and a prerequisite for investigating a targeted intervention for rehabilitation or prevention of falls involving hip abductor strength. Studies should be developed to assess on how progressive long-term hip abductor strength training affects fall rate reduction in order to implement new and tailored exercise programs. If a specific hip-abductor-strengthening intervention were shown to be effective for reducing fall risk, then we assume it would be easy for physiotherapists and other health care professionals to implement widely because the established ideas match their paradigm of the importance of hip abductor strength for mediolateral stability.

For the future development and possible implementation of fall-risk assessments utilizing hip abductor strength in daily clinical practice, clinicians might also be interested in assessing the change in patients' fall risk over time. To detect change over time, evaluative assessments must be used, and their responsiveness must

be assessed (17). Responsiveness indicates whether a measurement instrument can assess changes in health status over time (17) and should be investigated in a longitudinal study. If hip abductor strength is used to measure fall risk, then its responsiveness would be of interest for assessing change in fall risk over time after an intervention involving parameters other than hip abductor strength or when a person is followed over time, to evaluate changes in the risk of falls without intervention. Currently, we assessed the criterion validity of hip abductor strength, hence its diagnostic ability to discriminate between older fallers and non-fallers. In our systematic review we investigated the reliability and criterion validity of the included fall-risk assessments. Therefore, the present results are valid for diagnosing fallers versus non-fallers. If the studied instruments are used to evaluate change over time, additional information on their responsiveness and thus of their evaluation ability is needed.

Lastly, and as mentioned briefly in the section about the reference standard, near-falls could be an interesting topic for future research. Near-falls are clinically interesting but challenging in terms of recall bias (55). A near-fall could be defined as a slip, trip or loss of balance during which the person starts to fall but can stop or can avoid falling to the ground or a lower level (55). It was assumed that near-fallers are at a higher risk of falls than people with fewer near-falls in a community-dwelling setting (55). Therefore, recognizing near-fallers could be of great clinical interest, and future research should investigate the diagnostic accuracy of different fall-risk assessments and of hip abductor strength to differentiate between near-fallers and non-fallers.

CONCLUSION

This dissertation provides a new practical asset for authorities in the field of prevention, rehabilitation and movement science, allowing detecting and discriminating between older fallers and non-fallers. As a novel approach, the role of hip abductor strength in detecting the risk of falls of persons aged ≥ 65 was investigated. We evaluated the intra-rater reliability, SEM, and SDD of the hip-abductor-strength assessment in a side-lying and a standing position and assessed its criterion validity for detecting persons at risk of falls. We conclude that the assessment is reliable and feasible for clinical use and has good diagnostic accuracy to detect fall risk in this population. The assessment can be recommended for diagnostic purposes in daily clinical life. However, future research must assess its evaluative ability.

Given the probable necessity of combining hip abductor strength in a test battery with other tests for increased diagnostic ability, we performed a systematic review on the clinimetric properties of currently used fall-risk assessments. We found

different results for each setting, and the diagnostic accuracy of the investigated performance-based fall-risk assessments generally was not very high in clinical use. Overall, it was difficult to present firm conclusions because the sample sizes and number of the included studies were rather small for most of the settings.

This thesis also discusses critical points of the present and other studies as well as points that future studies should address. The benefits of this project concern a substantial percentage of people aged ≥ 65 because (over all settings) approximately 40% of the older population is at risk of falling. Therefore, positive social impacts and improved use of human resources can be expected.

REFERENCES

1. Vasunilashorn S, Steinman BA, Liebig PS, Pynoos J. Aging in place: evolution of a research topic whose time has come. *J Aging Res.* 2012;2012:120952.
2. United Nations DoEaSA, Population Division (2019). United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019: Highlights* (ST/ESA/SER.A/423).
3. Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC geriatrics.* 2011;11:33.
4. Liu SW, Obermeyer Z, Chang Y, Shankar KN. Frequency of ED revisits and death among older adults after a fall. *The American journal of emergency medicine.* 2015;33(8):1012-8.
5. Gill TM, Murphy TE, Gahbauer EA, Allore HG. Association of Injurious Falls With Disability Outcomes and Nursing Home Admissions in Community-Living Older Persons. *American Journal of Epidemiology.* 2013;178(3):418-25.
6. Sousa LM, Marques-Vieira CM, Caldevilla MN, Henriques CM, Severino SS, Caldeira SM. Risk for falls among community-dwelling older people: systematic literature review. *Revista gaucha de enfermagem.* 2017;37(4):e55030.
7. Gafner SC, Bastiaenen CH, Ferrari S, Gold G, Terrier P, Hilfiker R, et al. Hip muscle and hand-grip strength to differentiate between older fallers and non-fallers: a cross-sectional validity study. *Clinical interventions in aging.* 2018;13:1-8.
8. Gafner S, Bastiaenen CHG, Terrier P, Punt I, Ferrari S, Gold G, et al. Evaluation of hip abductor and adductor strength in the elderly: a reliability study. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity.* 2017;14:5.
9. Bruyneel AV, Gafner SC, Ferrari S, Gold G, Monnin D, Terrier P, et al. Intra-rater reliability of hip abductor isometric strength testing in a standing position in older fallers and non-fallers. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity.* 2018;15:9.
10. Hilliard MJ, Martinez KM, Janssen I, Edwards B, Mille ML, Zhang Y, et al. Lateral balance factors predict future falls in community-living older adults. *Archives of physical medicine and rehabilitation.* 2008;89(9):1708-13.
11. Arvin M, van Dieen JH, Faber GS, Pijnappels M, Hoozemans MJ, Verschueren SM. Hip abductor neuromuscular capacity: A limiting factor in mediolateral balance control in older adults? *Clinical biomechanics (Bristol, Avon).* 2016;37:27-33.
12. Arvin M, Hoozemans MJ, Burger BJ, Rispens SM, Verschueren SM, van Dieen JH, et al. Effects of hip abductor muscle fatigue on gait control and hip position sense in healthy older adults. *Gait & posture.* 2015;42(4):545-9.
13. Allet L, Kim H, Ashton-Miller J, De Mott T, Richardson JK. Frontal plane hip and ankle sensorimotor function, not age, predicts unipedal stance time. *Muscle & nerve.* 2012;45(4):578-85.
14. Daun F, Kibele A. Different strength declines in leg primary movers versus stabilizers across age-Implications for the risk of falls in older adults? *PloS one.* 2019;14(3):e0213361.
15. Widler KS, Glatthorn JF, Bizzini M, Impellizzeri FM, Munzinger U, Leunig M, et al. Assessment of hip abductor muscle strength. A validity and reliability study. *J Bone Joint Surg Am.* 2009;91(11):2666-72.
16. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *Journal of clinical epidemiology.* 2006;59(10):1033-9.
17. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology)*: Cambridge: Cambridge University Press.; 2011.
18. 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. *Journal of clinical epidemiology.* 2010;63(7):737-45.
19. 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. *Quality of Life Research.* 2018;27(5):1171-9.
20. 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. *Quality of Life Research.* 2018;27(5):1147-57.
21. Pijnappels M, van der Burg PJ, Reeves ND, van Dieen JH. Identification of elderly fallers by muscle strength measures. *European journal of applied physiology.* 2008;102(5):585-92.
22. Gafner SC, Bastiaenen CHG, Ferrari S, Gold G, Trombetti A, Terrier P, et al. The Role of Hip Abductor Strength in Identifying Older Persons at Risk of Falls: A Diagnostic Accuracy Study. *Clinical interventions in aging.* 2020;15:645-54.
23. Thapa PB, Brockman KG, Gideon P, Fought RL, Ray WA. Injurious falls in nonambulatory nursing home residents: a comparative study of circumstances, incidence, and risk factors. *Journal of the American Geriatrics Society.* 1996;44(3):273-8.
24. Schoene D, Wu SM, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, et al. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. *Journal of the American Geriatrics Society.* 2013;61(2):202-8.
25. Nelson AJ, Certo LJ, Lembo LS, Lopez DA, Manfredonia EF, Vanichpong SK, et al. The functional ambulation performance of elderly fallers and non-fallers walking at their preferred velocity. *NeuroRehabilitation.* 1999;13(3):141-6.

26. Johnson ME, Mille ML, Martinez KM, Crombie G, Rogers MW. Age-related changes in hip abductor and adductor joint torques. *Archives of physical medicine and rehabilitation*. 2004;85(4):593-7.
27. Gafner SC, Bastiaenen CHG, Biver E, Ferrari S, Allet L. Reliability and validity of an adapted hip abductor strength measure as a potential new fall risk assessment for older persons: a study protocol. *BMC geriatrics*. 2021;21(1):110.
28. World Health O. International classification of functioning, disability and health : ICF. Geneva: World Health Organization; 2001.
29. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ (Clinical research ed)*. 2015;351:h5527.
30. WHO World Health Organization report/ WHO Global report on falls prevention in older age. 2007. ISBN 978 92 4 156353 6, 2007.
31. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine*. 2000;45(1-2):23-41.
32. Santos GM, Souza AC, Virtuoso JF, Tavares GM, Mazo GZ. Predictive values at risk of falling in physically active and no active elderly with Berg Balance Scale. *Rev Bras Fisioter*. 2011;15(2):95-101.
33. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E, et al. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. *Journal of geriatric physical therapy* (2001). 2017;40(1):1-36.
34. Gordis L. *Epidemiology*. Philadelphia: Elsevier/Saunders; 2009.
35. Tinetti ME, Kumar C. The patient who falls: "It's always a trade-off". *Jama*. 2010;303(3):258-66.
36. Park SH. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. *Aging clinical and experimental research*. 2018;30(1):1-16.
37. Nunan S, Brown Wilson C, Henwood T, Parker D. Fall risk assessment tools for use among older adults in long-term care settings: A systematic review of the literature. *Australasian journal on ageing*. 2018;37(1):23-33.
38. Ganz DA, Higashi T, Rubenstein LZ. Monitoring Falls in Cohort Studies of Community-Dwelling Older People: Effect of the Recall Interval. *Journal of the American Geriatrics Society*. 2005;53(12):2190-4.
39. Cummings SR, Nevitt MC, Kidd S. Forgetting falls. The limited accuracy of recall of falls in the elderly. *Journal of the American Geriatrics Society*. 1988;36(7):613-6.
40. Hoffman GJ, Ha J, Alexander NB, Langa KM, Tinetti M, Min LC. Underreporting of Fall Injuries of Older Adults: Implications for Wellness Visit Fall Risk Screening. *Journal of the American Geriatrics Society*. 2018;66(6):1195-200.
41. Gafner SC, Allet L, Hilfiker R, Bastiaenen CHG. Reliability and diagnostic accuracy of commonly used performance tests relative to fall history in older persons: a systematic review Clinical interventions in aging. 2021;Under review.
42. Faber MJ, Bosscher RJ, van Wieringen PC. Clinimetric properties of the performance-oriented mobility assessment. *Physical therapy*. 2006;86(7):944-54.
43. Hars M, Audet MC, Herrmann F, De Chasse J, Rizzoli R, Reny JL, et al. Functional Performances on Admission Predict In-hospital Falls, Injurious Falls and Fractures in Older Patients: a Prospective Study. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2018.
44. Viveiro LAP, Gomes GCV, Bacha JMR, Carvas Junior N, Kallas ME, Reis M, et al. Reliability, Validity, and Ability to Identify Fall Status of the Berg Balance Scale, Balance Evaluation Systems Test (BESTest), Mini-BESTest, and Brief-BESTest in Older Adults Who Live in Nursing Homes. *Journal of geriatric physical therapy* (2001). 2019;42(4):E45-e54.
45. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Physical therapy*. 2000;80(9):896-903.
46. Muir SW, Berg K, Chesworth B, Speechley M. Use of the Berg Balance Scale for Predicting Multiple Falls in Community-Dwelling Elderly People: A Prospective Study. *Physical therapy*. 2008;88(4):449-59.
47. Bento PC, Pereira G, Ugrinowitsch C, Rodacki AL. Peak torque and rate of torque development in elderly with and without fall history. *Clinical biomechanics (Bristol, Avon)*. 2010;25(5):450-4.
48. Maffiuletti NA, Aagaard P, Blazevich AJ, Folland J, Tillin N, Duchateau J. Rate of force development: physiological and methodological considerations. *European journal of applied physiology*. 2016;116(6):1091-116.
49. Seidler RD, Bernard JA, Burutolu TB, Fling BW, Gordon MT, Gwin JT, et al. Motor control and aging: Links to age-related brain structural, functional, and biochemical effects. *Neuroscience and Biobehavioral Reviews*. 2010;34(5):721-33.
50. Morcelli MH, Rossi DM, Karuka AH, Crozara LF, Hallal CZ, Marques NR, et al. Peak torque, reaction time, and rate of torque development of hip abductors and adductors of older women. *Physiotherapy theory and practice*. 2016;32(1):45-52.
51. Andersen LL, Aagaard P. Influence of maximal muscle strength and intrinsic muscle contractile properties on contractile rate of force development. *European journal of applied physiology*. 2006;96(1):46-52.
52. Panel on Prevention of Falls in Older Persons AGS, British Geriatrics S. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. *Journal of the American Geriatrics Society*. 2011;59(1):148-57.
53. Stevens JA, Phelan EA. Development of STEADI: A Fall Prevention Resource for Health Care Providers. *Health Promotion Practice*. 2013;14(5):706-14.
54. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016;16(1):97.
55. Arnold CM, Faulkner RA. The history of falls and the association of the timed up and go test to falls and near-falls in older adults with hip osteoarthritis. *BMC geriatrics*. 2007;7:17.

CHAPTER 9

Impact Paragraph

At the end of this PhD trajectory, it is important for me to reflect on the scientific and societal significance of my research projects and, most importantly, on their significance at the patient level. The present research was aimed at enhancing clinical work and improving patients' outcomes by identifying relevant parameters for the future development of a targeted therapy for fall risk.

The results of this thesis indicate that assessing hip abductor strength helps with understanding older persons' fall risks. The hip-abductor-strength assessment is feasible in older persons, shows good clinimetric properties and provides additional value to currently used fall risk evaluations and, thus, to fall risk prevention and rehabilitation programs. It evaluates complementary aspects of falls but is not yet measured systematically. Furthermore, currently used performance-based fall-risk assessments only show moderate ability to discriminate between fallers and non-fallers, and the setting where the tests will be performed strongly influences which tests can be recommended for clinical use.

Scientific and social relevance of this research

Regarding the scientific relevance, this is the first project to investigate all hip muscle strength groups in the same study, subsequently confirming our hypotheses on the importance of hip abductor strength in discriminating between fallers and non-fallers. This is a highly relevant result considering the growing proportion of persons aged ≥ 65 years and the related need for actions to improve fall-risk detection and prevention. The systematic review showed that currently used fall-risk assessments only have moderate diagnostic accuracy, that single performance tests are less investigated than multidimensional fall-risk assessments and that the often-used multidimensional fall-risk assessments do not necessarily have a better ability to discriminate between fallers and non-fallers than single performance tests do. Regarding the good diagnostic accuracy of hip abductor strength, single performance tests should be investigated in greater detail, especially given that they usually do not take long to perform and give a good indication of what treatment should be targeted. For future research, this knowledge about the importance of hip abductor strength should encourage investigations into effective interventions with which to decrease the risk of falls among persons aged ≥ 65 years.

To increase the scientific impact, the results of this thesis have been presented at national and international conferences and published in scientific journals.

Due to the direct clinical applicability of our findings, the results are of interest for scientists, clinicians and teachers, for whom the relationship between hip abductor strength and mediolateral stability seems evident. The clinical value and scientific validation of the present findings facilitate their implementation into

clinical practice. To close the gap between research and practice, we presented the results by healthcare institution (e.g., at hospitals and nursing homes). In addition, we have started to integrate the results in continuing education classes on geriatrics at regional physiotherapy associations in Switzerland and at the University of Applied Sciences and Arts of Western Switzerland.

The interest of physiotherapists in our results and our exchanges with different clinicians encouraged us to evolve the measurement instrument and method. We aimed to develop a measurement procedure to measure hip abductor muscle strength in a functional standing and a closed chain position, using an instrument that is easy to transport and manipulate, and that should directly display the measured values without needing time-consuming post hoc data extraction. These are essential issues with which to guarantee the wide clinical use of our measurement instrument and thus to create a long-term impact on healthcare.

A prototype considering these needs was developed in collaboration with engineers at the Institute of Industrial Technologies and Sciences, University of Applied Sciences and Arts Western Switzerland. Funding applications were submitted for further development of our invention and to assess the reliability (inter and intra-rater) and criterion validity of the newly developed measure and procedure. (A multicenter study is currently ongoing.) An agreement was signed among the University of Geneva, the University Hospitals of Geneva and the University of Applied Sciences and Arts of Western Switzerland for the intellectual property as well as for first funding for patent submission. The patent for the new measurement instrument and procedure was filed at the European Patent Office in October 2020 (application and patent number 202000 13.9-1115). Discussions are currently underway with potential companies and commercial partners to develop the prototype further, submit the final patent and commercialize our measurement instrument. The commercialization of this new test procedure will be an important step for implementing hip abductor strength as an indicator of fall risk and adding it to a well-targeted prevention or rehabilitation intervention.

At the patient level, this dissertation contributes important information on a new, easy to use fall-risk assessment. Moreover, the explanations provided to assist clinicians in understanding how to use and interpret the diagnostic accuracy of the measurement instruments should improve the use of fall-risk assessments and result in earlier detection of older persons at risk of falls. The quality of life of the concerned persons will improve (or stay constant) because they will be able to stay independent for as long as possible during their daily life activities in their known surroundings and participate in social activities.

Socially, we underline that early detection of persons at risk of falls allows for better prevention of first and/or recurrent falls, which will impact societal and

economic concerns and enhance the use of human resources because treatments can be targeted better. In this thesis, we did not investigate the economic impact of improved fall-risk detection. However, falls have significant medical, societal and financial consequences and contribute to the huge economic costs for aging societies.

Valorization of the results for specific target groups and their perspectives of the future

The most important group benefitting from the present results are patients aged ≥ 65 years. Both today and in the future, better-targeted fall-risk detection and tests with good diagnostic accuracy will help to recognize persons susceptible to falls and assist them in choosing interventions to decrease this risk. Individual goal setting by the patient, together with the healthcare professional, will help patients to maintain their actual state of independence and take an active part in their familiar and societal environment for as long as possible. Many organizations grouping older persons exist that can offer a wide variety of activities and courses. Knowledge among the coaches and leaders of the groups will allow the implementation and transfer of knowledge from healthcare professionals to persons aged ≥ 65 years and thereafter within the population of older persons itself.

Healthcare professionals are another large group that will benefit from the results of this dissertation. Better knowledge of the diagnostic accuracy of newly investigated and currently used fall-risk assessments will allow for well-targeted detection of persons at risk as well as interventions. However, in the future, it will be necessary to transfer the research findings to health care professionals in order to create inside knowledge that will allow caregivers to correctly interpret the test's diagnostic accuracy and results. This knowledge transfer will reinforce the clinical reasoning process together with the patient and help healthcare professionals to use the test results as a probability of fall risk, rather than a dichotomous yes or no value. The systematic review shows that fall risk detection should be a process between the health care professional and the concerned person that considers the results of several fall-risk assessment tests, the person's setting and personal goal setting and other information.

Researchers are another group benefitting from the insights the different projects described in this thesis. Our results are a nice starting point for future research projects investigating possible interventions including hip abductor strength and for further research on improving the ability to detect persons at risk of falls (with a focus on single but also multidimensional performance tests).

Therefore, the benefits of this dissertation concern a substantial percentage of persons aged ≥ 65 years, given that (over all settings) approximately 40% of the older population is at risk of falling, as well as healthcare professionals, researchers and thus society.

CONCLUSION

The knowledge gained from the different projects comprising this PhD thesis will be beneficial for researchers, healthcare professionals, persons aged ≥ 65 years, caregivers and relatives of people who are at risk of falls. Therefore, it will have high scientific, societal and clinical impacts. We recommend assessing hip abductor strength for diagnostic purposes and encourage its assessment in daily clinical use.

ADDENDUM

Summary

Zusammenfassung

Résumé

Acknowledgments

About the author

SUMMARY

This thesis aimed to gain insight into the role of hip abductor muscle strength in fall risk in older adults and assess its ability to discriminate between older fallers and non-fallers. We were interested in the possible potential of hip abductor strength as a fall-risk assessment either in single use or within a test battery.

The **Peface** gives an introduction in a scenario that we all lived and makes us realize the importance of falls and their associated deleterious consequences. The preface allows an emotional entry into research, which underlines that the present project is not just « one more of those projects about fall risk assessments... ». The preface makes us realize how much this subject can touch us personally and how much we are emotionally linked to it.

- What would you like to know to support your loved ones in preventing a fall?
- Which questions would you like to have answered in relation to fall risk?
 - o Which factors influence falls?
 - o What tests can evaluate those fall-influencing factors?
 - o What does science say regarding fall risk and fall-risk assessments?

Chapter 1 introduces the importance of the theme of fall risk by underlying the demographic changes that are leading to a drastically increasing number of older people, which implicates an associated increasing fall incidence. Many factors influence the construct “fall risk”. However, the focus of this dissertation is on the physiological factors, the parameters that are modifiable and important for guiding the content of a physical therapy intervention, such as decreased balance, strength, and mobility. Chapter 1 further discusses the necessary aspects we need to understand such as falls and fall-influencing factors, what kind of test do exist for evaluating those fall-influencing factors, the influence of lower limb muscle strength and more specifically the reasoning of the interest in hip abductor strength in relation to fall risk. As such, Chapter 1 gives information about the important parameters for the choice of a test for fall-risk assessment and the taxonomy, definitions, and methodological aspects of clinimetric studies.

In fact, functional lower-extremity weakness in older persons is a strong indicator that could enhance the unfortunate transition from a non-faller into a faller, and its assessment and treatment for persons at risk of falls is recommended. Recent literature underlines the significance of hip abductor muscle strength to maintain mediolateral balance control, a factor that is often reduced in older persons, especially fallers. Hip frontal plane muscle strength and spinal moments of force are the dominant muscles to counteract destabilization in the mediolateral direction

during single leg support and showed a more pronounced decline in strength compared to muscles acting in the sagittal plane. The mentioned facts led to the hypothesis that hip abductor strength could be a promising contributing factor for assessing fall risk. Because this approach seemed to be clinically relevant and novel, we decided to investigate the hip abductor muscle group in greater detail. We planned to establish cut-off values and diagnostic accuracy to facilitate the clinical decision-making process regarding older individuals' fall risks. The validation of this approach provides clinicians not only an additional fall-risk assessment to the already existing test battery but might also give crucial information about a directly targetable parameter for fall prevention.

With this dissertation, we addressed two main issues, which are divided in Part 1 and Part 2.

Part 1: The role of hip abductor strength in the field of fall-risk assessment and development of a new measurement instrument and procedure for daily clinical use

The first part of this dissertation (Chapters 2 to 6) concentrated on the role of hip abductor strength in the assessment of fall risk of persons 65 years and over. It was planned to assess feasibility, reliability, and validity of our measurement method for hip frontal plane strength assessments. In a later emerging project, we further develop a hip abductor strength test to obtain a new and functional testing procedure that is easy to use.

Within **Chapters 2 and 3**, the feasibility and intra-rater reliability of the hip frontal plane strength assessment in side-lying position and the hip abductor strength assessment in a standing position were assessed. Relative and absolute reliability are a prerequisite for good validity. In **Chapter 2** we used a standardized way to measure hip abductor and adductor strength in which the participant was in a side-lying position and pushed his or her leg as hard and fast as possible against a dynamometer fixed to a custom-made, metallic frame that was fixed to the treatment table. Seventy-six participants aged 65 years and over were recruited from a geriatric hospital that is part of the Geneva University Hospitals and in an outpatient practice in Switzerland. Thirty-eight persons underwent the abduction and the other 38 persons the adduction test. The participants were tested twice within one to three hours by the same rater. Intraclass correlation coefficient ($ICC_{\text{agreement}}$), standard error of measurement (SEM), and smallest detectable difference (SDD) were determined and showed that hip abductor and adductor maximum voluntary isometric strength (MVIS) and rate of force generation (RFG) assessment in side-lying position is feasible and shows high values for reliability in persons 65 and older.

However, the assessment time of around 10min in side-lying position is too long for daily clinical use.

This led to the assessment of the feasibility and reliability of the hip abductor strength assessment in a standing position in **Chapter 3**. As for Chapter 2 individuals over 65 years of age were recruited from the geriatric department of a university hospital and an outpatient clinic. Thirty-two older subjects, including 16 fallers (≥ 1 fall during the last 12 months) and 16 non-fallers were included. MVIS and RFG of the hip abductors of the right leg were evaluated in a standing position using a hand-held dynamometer. Two test-sessions were carried out by the same rater one to three hours apart. Reliability was determined using the ICC_{agreement}, the SEM and a Bland and Altman analysis. The results showed that hip abductor strength assessment in a standing position is feasible and achieves ICC_{agreement} values that are as high as the ones assessed in a side-lying position. The advantage of the test in standing position is that it takes about half the time the hip abductor strength assessment takes in a side-lying position and hence seems much more feasible in clinical practice. Although both testing positions showed a good intra-rater reliability, in both cases, high SEM and SDD were evaluated, with values of SDD that are slightly lower in the standing than in the side-lying position.

To be able to answer the question about the importance of hip abductor strength in relation to fall risk, we evaluated in **Chapter 4** which of the six hip muscle groups best discriminates between older fallers and non-fallers and compared their discriminative ability with the one of the handgrip strengths. In 60 older persons aged 65 years and over (38 females, 22 males) the MVIS and RFG of hip abductors, adductors, internal and external rotators, extensors and flexors were measured with a dynamometer fixed to a custom-made frame as well as hand-grip strength with a Martin Vigorimeter. It turned out that hip abductor strength was the only assessed hip muscle strength group that showed a discriminative ability that was predefined as good enough for clinical use ($AUC > 0.7$), as compared to the chosen reference standard fall history. Therefore, the results of Chapter 4 indicate that hip abductor MVIS is a useful measure to distinguish between older adult fallers and non-fallers regarding the chosen external criterion.

This led us to continue with the evaluation of the specific psychometric properties of the hip abductor strength in **Chapter 5**. Chapter 5 is an extended evaluation and a next step to the results of Chapter 4. The diagnostic accuracy of hip abductor muscle strength was examined to decide whether this parameter should be integrated in a test battery to assess older persons' fall risks. Thus, we also calculated net sensitivity and specificity to get insight in its possible combination with our chosen fall risk assessments, short physical performance battery (SPPB) and Timed up and Go (TUG) test. All of the 60 included patients' demographic data

(e.g. age, sex, and BMI) were extracted from the clinical database, the reference standard fall history was collected by verbal questioning and thereafter, trained physical therapists (blinded to the participants' fall histories) assessed the Falls Efficacy Scale International, Mini Mental State Examination, the SPPB and TUG, as well as participants' hip abductor strength (MVIS and RFG). The results of Chapter 5 indicate that hip abductor strength (MVIS and RFG) shows good diagnostic accuracy to differentiate between older fallers and non-fallers. The high net sensitivity when hip ABD strength measurements (MVIS and RFG) are combined with currently used fall risk assessments shows its promising contributing value to a test battery. Hip abductor strength is an easily measurable parameter and should, on a daily clinically basis, be considered as an important influencing factor of fall risk.

However, during the different projects of Chapters 2 to 5, we were confronted with more challenges in relation to the assessment position as well as the measurement instrument for hip abductor strength assessment and its functional and mostly clinical applicability throughout the mentioned studies. This led us to reflect, and inspired a way to assess hip abductor strength in daily clinical life. Therefore, with the focus on allowing our previous findings to be integrated in clinical practice, the additional project for the development of a measurement procedure in standing position emerged for this thesis and is addressed in **Chapter 6**. With the assistance of engineers of the Institute of Industrial Technologies and Sciences, University of Applied Sciences and Arts Western Switzerland (HES-SO Geneva/HEPIA), a prototype of an easy to manipulate and transportable measurement instrument and associated measurement procedure was developed and a patent was filed in October 2020 (impact paragraph). Consequently, with the goal of evaluating the adapted measurement procedure to allow a later use in daily clinical life, the study protocol of a reliability and validity study with the adapted measurement instrument and procedure was developed and is elaborated in Chapter 6.

Part 2: The diagnostic accuracy of currently used fall-risk assessments (including strength measures) to gain insight into their ability for fall risk detection and/or prediction

Chapter 7 assessed in a systematic review the clinimetric properties of other tests that assess body functions and structures and/or the ICF "activities and participation" components in relation to fall risk of older persons. Aware of the multifactorial etiology of falls and assuming that hip abductor strength will have to be combined with other performance-based fall risk assessments for a good fall risk detection, we aimed to provide an overview of the reliability (inter- and intra-rater) and diagnostic accuracy, specifically post-test probability, of commonly used performance tests to assess the

fall risk of persons 65 years and older, according to their setting (community dwelling, nursing home, or hospital). The goal was that this overview can support caregivers in making decisions with the person at risk of falls to help them choose a suitable fall-risk assessment corresponding to the patient's specific setting.

Overall, 42 articles were retained for the systematic review of which 21 articles contained information of reliability and 25 of validity. Four of the included articles contained information for reliability and validity of the included test. To facilitate the clinical applicability and because the diagnostic accuracy of tests is strongly influenced by the population under study, we summarized our findings according to the settings of the results (community-dwelling, nursing home, hospital, and combined setting of hospitalized and community-dwelling persons). In general, the systematic review showed that the discriminative ability of the included tests to detect persons at risk of falling is not very good for clinical use. Most of the included tests only show relatively low sensitivity values and small changes from pre- to post-test probabilities. Thus, our results as well as the knowledge of the multifactorial etiology of falls underline that the currently used fall risk assessments are in a single-use not good enough to discriminate between older fallers and non-fallers. To be able to choose an appropriate fall risk assessment health care professionals should know and understand the clinimetric properties of a test. In our systematic review, we explain important aspects of a measurement instrument, their influencing factors and their interpretation for clinicians. A clinical application of its results in everyday practice can easily be done and supports the decision-making process between patient and caregiver.

Chapter 8 and Chapter 9 are completing this thesis with a general discussion and the impact paragraph. **Chapter 8** discusses our main findings, presents reflections on strength, limitations and methodological aspects of our studies, and provides implications for clinical practice and future research. Assessing hip abductor strength helps with understanding older persons' fall risks. The hip-abductor-strength assessment is feasible in older persons, shows good clinimetric properties and provides additional value to currently used fall risk evaluations and, thus, to fall risk prevention and rehabilitation programs. It evaluates complementary aspects of falls but is not yet measured systematically. Furthermore, currently used performance-based fall-risk assessments only show moderate ability to discriminate between fallers and non-fallers, and the setting where the tests will be performed strongly influences which tests can be recommended for clinical use. The clinical applicability of the findings is highlighted in Chapter 8. Importance is given to the fact that researchers must make their work understandable for direct application to clinical life and that closing the gap between research and practice to develop best-practice guidelines for patients is required.

Finally, **Chapter 9** reflects on the scientific and societal significance of my research projects and, most importantly, the significance of the results at the patient level. The present research aimed to enhance clinical work and improving patients' outcomes by identifying relevant parameters for the future development of a targeted therapy for fall risk. The knowledge gained from the different projects comprising this PhD thesis are beneficial for researchers, healthcare professionals, persons aged ≥ 65 years, and relatives of people who are at risk of falls. Therefore, it has high scientific, societal and clinical impacts. We recommend assessing hip abductor strength for diagnostic purposes and encourage its assessment in daily clinical use.

ZUSAMMENFASSUNG

Ziel dieser Arbeit war es, einen Einblick in die Rolle der Hüftabduktoren-Kraft bezüglich des Sturzrisikos älterer Menschen zu gewinnen und zu bewerten, ob mittels Hüftabduktoren-Kraft zwischen älteren Personen, die stürzen, und solchen, die nicht stürzen, unterschieden werden kann. Wir interessierten uns für das mögliche Potenzial der Kraft der Hüftabduktoren zur Sturzrisikobestimmung als Einzeltest oder innerhalb einer Testbatterie.

Das **Vorwort** gibt eine Einführung in ein Szenario, das wir alle erlebt haben, und macht uns die Bedeutung von Stürzen und die damit verbundenen negativen Folgen bewusst. Es ermöglicht einen emotionalen Einstieg in die Forschung, der unterstreicht, dass es sich bei dem vorliegenden Projekt nicht nur um „ein weiteres dieser Projekte zur Bewertung von Sturzrisiken“ handelt. Das Vorwort zeigt in dieser Weise auf, wie sehr uns dieses Thema persönlich berühren kann und wie sehr wir emotional damit verbunden sind.

- Was möchten Sie wissen, um Ihre Angehörigen bei der Vorbeugung eines Sturzes zu unterstützen?
- Welche Fragen würden Sie gerne in Bezug auf das Sturzrisiko beantwortet haben?
 - o *Welche Faktoren beeinflussen Stürze?*
 - o *Welche Tests können diese sturzbeeinflussenden Faktoren beurteilen?*
 - o *Was sagt die Wissenschaft zum Sturzrisiko und zur Sturzrisikobewertung?*

Kapitel 1 führt in die Bedeutung des Themas „Sturzrisiko“ ein, indem es die demografischen Veränderungen erläutert, die zu einer drastisch steigenden Anzahl älterer Menschen führen, was wiederum eine damit verbundene steigende Sturzhäufigkeit impliziert. Viele Faktoren beeinflussen das Konstrukt "Sturzrisiko". Der Schwerpunkt dieser Arbeit liegt jedoch auf den Parametern, die veränderbar und für die inhaltliche Ausrichtung einer physiotherapeutischen Intervention wichtig sind, wie z. B. vermindertes Gleichgewicht, Kraft und Mobilität. In Kapitel 1 werden die notwendigen Aspekte erörtert, die wir verstehen müssen, wie z. B. Stürze und sturzbeeinflussende Faktoren, welche Art von Tests es zur Bewertung dieser sturzbeeinflussenden Faktoren gibt, der Einfluss der Muskelkraft der unteren Extremität und insbesondere die Begründung für das Interesse an der Kraft der Hüftabduktoren im Zusammenhang mit dem Sturzrisiko. In diesem Sinne klärt Kapitel 1 über die wichtigen Parameter für die Wahl eines Tests zur Bewertung des Sturzrisikos sowie über die Taxonomie, die Definitionen und die methodischen Aspekte von klinimetrischen Studien auf.

Tatsächlich ist eine funktionelle Schwäche der unteren Extremitäten bei älteren Menschen ein deutlicher Indikator, der den unglücklichen Übergang von einer Person ohne Sturzrisiko zu einer stürzenden Person begünstigen könnte. Die Bewertung der Kraft der unteren Extremität und deren Behandlung wird deshalb bei sturzgefährdeten Personen empfohlen. In der neueren Literatur wird die Bedeutung der Kraft der Hüftabduktoren für die Aufrechterhaltung der mediolateralen Gleichgewichtskontrolle hervorgehoben und ist ein Faktor, der bei älteren Menschen, insbesondere bei sturzgefährdeten Personen, häufig vermindert ist. Die Kraft der in der Frontalebene dominanten Hüftabduktoren wirkt einer Destabilisierung in mediolateraler Richtung während des Einbeinstandes entgegen und zeigte einen stärkeren Kraftabfall im Vergleich zu den in der Sagittalebene wirkenden Muskeln. Die genannten Fakten führten zu der Hypothese, dass die Kraft der Hüftabduktoren ein vielversprechender Faktor für die Bewertung des Sturzrisikos sein könnte. Da dieser Ansatz klinisch relevant und neu zu sein schien, beschlossen wir, die Muskelgruppe der Hüftabduktoren genauer zu untersuchen. Wir planten, Schwellenwerte und diagnostische Genauigkeit festzulegen, um den klinischen Entscheidungsprozess hinsichtlich des Sturzrisikos älterer Menschen zu erleichtern. Die Validierung dieses Ansatzes bietet Kliniker*innen nicht nur eine zusätzliche Bewertung des Sturzrisikos zur bereits bestehenden Testbatterie, sondern könnte auch entscheidende Informationen über einen direkt beeinflussbaren Parameter für die Sturzprävention liefern.

In dieser Dissertation haben wir zwei Hauptthemen behandelt, die in Teil 1 und Teil 2 unterteilt sind.

Teil 1: Die Rolle der Kraft der Hüftabduktoren im Bereich der Sturzrisikobewertung und die Entwicklung eines neuen Messinstruments und Messverfahrens für den klinischen Alltag

Der erste Teil dieser Dissertation (Kapitel 2 bis 6) konzentrierte sich auf die Rolle der Kraft der Hüftabduktoren bei der Beurteilung des Sturzrisikos von Personen über 65 Jahren. Es war geplant, die Durchführbarkeit, Reliabilität und Validität unserer Messmethode zur Beurteilung der Kraft der Hüftabduktoren in der Frontalebene zu untersuchen. In einem späteren Projekt entwickeln wir den Krafttest für die Hüftabduktoren weiter, um ein neues und funktionelles Testverfahren zu erhalten, das einfach zu handhaben ist für den Gebrauch in der Praxis.

In den **Kapiteln 2 und 3** wurden die Durchführbarkeit und die Intra-Rater-Reliabilität der Messung der Kraft in der Hüfte in der Frontalebene in Seitenlage und der Bewertung der Kraft der Hüftabduktoren in stehender Position bewertet. Relative und absolute Reliabilität sind eine Voraussetzung für eine gute Validität. In

Kapitel 2 verwendeten wir eine standardisierte Methode zur Messung der Kraft der Hüftabduktoren und -adduktoren. Dazu mussten die Teilnehmenden in Seitenlage ein Bein so stark und schnell wie möglich gegen einen Dynamometer drücken, welcher an einem speziell angefertigten Metallrahmen fixiert und der wiederum auf dem Behandlungstisch befestigt war. Sechundsiebzig Teilnehmende im Alter von 65 Jahren und älter wurden in einem geriatrischen Krankenhaus der Genfer Universitätskliniken und in einer ambulanten Praxis in der Schweiz rekrutiert. Achtunddreissig Personen unterzogen sich dem Abduktionstest und die anderen 38 Personen dem Adduktionstest. Die Teilnehmenden wurden zweimal innerhalb von ein bis drei Stunden von derselben Testperson getestet. Der Intraclass-Korrelationskoeffizient ($ICC_{\text{agreement}}$), der Standardmessfehler (SEM) und die kleinste nachweisbare Differenz (SDD) wurden bestimmt und zeigten, dass die Bewertung der maximalen, willkürlichen isometrischen Kraft der Hüftabduktoren und -adduktoren (Maximum Voluntary Isometric Strength (MVIS)) und der Krafterzeugungsrate (Rate of Force Generation (RFG)) in Seitenlage durchführbar ist und hohe Werte für die Reliabilität bei Personen ab 65 Jahren aufweist. Allerdings ist die Messzeit von etwa 10 Minuten in Seitenlage für den täglichen klinischen Einsatz zu lang.

Dies führte in **Kapitel 3** zur weiteren Beurteilung der Durchführbarkeit und Reliabilität der Bewertung der Hüftabduktorenkraft im Stehen. Wie in Kapitel 2 wurden Personen im Alter von über 65 Jahren aus der geriatrischen Abteilung des Genfer Universitätsspitals und einer ambulanten Praxis rekrutiert. Zweiunddreißig ältere Probanden, darunter 16 Sturzpatient*innen (≥ 1 Sturz in den letzten 12 Monaten) und 16 Personen, die nicht gestürzt sind, wurden eingeschlossen. MVIS und RFG der Hüftabduktoren des rechten Beins wurden in stehender Position mit einem Handdynamometer gemessen. Zwei Testsitzungen wurden im Abstand von ein bis drei Stunden von derselben Untersuchungsperson durchgeführt. Die Reliabilität wurde anhand des $ICC_{\text{agreement}}$, des SEM und einer Bland-Altman-Analyse ermittelt. Die Ergebnisse zeigten, dass die Messung der Kraft der Hüftabduktoren im Stehen durchführbar ist und dass die Werte für den $ICC_{\text{agreement}}$ genauso hoch sind wie die Werte, die in Seitenlage ermittelt wurden. Der Vorteil des Tests im Stehen besteht darin, dass er etwa die Hälfte der Zeit in Anspruch nimmt, die für die Messung der Kraft der Hüftabduktoren in Seitenlage benötigt wird, und daher in der klinischen Praxis viel praktikabler erscheint. Obwohl beide Testpositionen eine gute Intra-Rater-Reliabilität aufwiesen, wurden in beiden Fällen hohe SEM- und SDD-Werte ermittelt, wobei die SDD-Werte in der stehenden Position etwas niedriger waren als in der Seitenlage.

Um die Frage nach der Bedeutung der Hüftabduktoren-Kraft im Zusammenhang mit dem Sturzrisiko beantworten zu können, haben wir in **Kapitel 4** untersucht, welche der sechs Hüftmuskelgruppen am besten zwischen älteren

Personen die stürzen und Personen die nicht stürzen unterscheidet. Zudem habe wir ihre Unterscheidungsfähigkeit mit derjenigen der Handgriffkraft verglichen. Bei 60 älteren Personen im Alter von 65 Jahren und älter (38 Frauen, 22 Männer) wurden die MVIS und RFG der Hüftabduktoren, Adduktoren, Innen- und Aussenrotatoren, Extensoren und Flexoren mit einem an einem speziell angefertigten Rahmen befestigten Dynamometer sowie die Handgriffstärke mit einem Martin Vigorimeter gemessen. Es stellte sich heraus, dass die Kraft der Hüftabduktoren die einzige Hüftmuskelkraftgruppe war, die eine für den klinischen Gebrauch ausreichende Unterscheidungsfähigkeit (Area Under the Curve (AUC) $> 0,7$) im Vergleich zum gewählten Referenzstandard „Anzahl Stürze im vergangenen Jahr“ aufwies. Daher zeigen die Ergebnisse von Kapitel 4, dass die Kraft der Hüftabduktoren ein nützliches Messinstrument zur Unterscheidung zwischen älteren Erwachsenen, die stürzen, und solchen, die nicht stürzen, in Bezug auf das gewählte externe Kriterium ist („Anzahl Stürze im vergangenen Jahr“).

Dies veranlasste uns, in **Kapitel 5** mit der Auswertung der spezifischen psychometrischen Eigenschaften der Kraft der Hüftabduktoren fortzufahren. Kapitel 5 ist eine erweiterte Auswertung und schliesst an die Ergebnisse von Kapitel 4 an. Die diagnostische Genauigkeit der Kraft der Hüftabduktoren wurde untersucht, um zu entscheiden, ob dieser Parameter in eine Testbatterie zur Bewertung des Sturzrisikos älterer Menschen integriert werden sollte. Wir berechneten deshalb auch die Nettosensitivität und -spezifität, um einen Einblick in die mögliche Kombination mit den von uns gewählten Sturzrisikotests, der Short Physical Performance Battery (SPPB) und dem Timed Up and Go (TUG)-Test, zu erhalten. Alle demografischen Daten der 60 eingeschlossenen Patient*innen (z. B. Alter, Geschlecht und BMI) wurden der klinischen Datenbank entnommen, der Referenzstandard „Sturzanamnese“ wurde durch mündliche Befragung erhoben und anschliessend haben geschulte Physiotherapeut*innen (die die Sturzanamnese der Teilnehmenden nicht kannten) die Falls Efficacy Scale International, die Mini Mental State Examination, die SPPB, den TUG sowie die Kraft der Hüftabduktoren der Teilnehmenden (MVIS und RFG) erfasst. Die Ergebnisse von Kapitel 5 zeigen, dass die Kraft der Hüftabduktoren (MVIS und RFG) eine gute diagnostische Genauigkeit aufweist, um zwischen älteren Sturzpatient*innen und Personen die nicht stürzen zu unterscheiden. Die hohe Nettosensitivität, wenn Messungen der Kraft der Hüftabduktoren (MVIS und RFG) mit den derzeit verwendeten Sturzrisikotests (SPPB und TUG) kombiniert werden, zeigt, dass sie einen vielversprechenden Beitrag zu einer Testbatterie leisten können. Die Kraft der Hüftabduktoren ist ein leicht messbarer Parameter und sollte im klinischen Alltag als ein wichtiger Einflussfaktor des Sturzrisikos betrachtet werden.

Während der verschiedenen Projekte in den Kapiteln 2 bis 5 sahen wir uns jedoch mit verschiedenen Herausforderungen in Bezug auf die Testpositionen

sowie dem Messinstrument für die Bewertung der Kraft der Hüftabduktoren und ihre funktionelle und vor allem klinische Anwendbarkeit in den genannten Studien konfrontiert. Dies veranlasste uns zum Nachdenken und inspirierte uns zur Entwicklung eines neuen Prozederes zur Messung der Kraft der Hüftabduktoren im klinischen Alltag. Mit dem Fokus, unsere bisherigen Erkenntnisse in die klinische Praxis einfließen zu lassen, entstand für diese These daher ein zusätzliches Projekt zur Entwicklung eines Messverfahrens der Kraft der Hüftabduktoren im Stehen, das **in Kapitel 6** beschrieben wird. Mit Unterstützung von Ingenieuren des Instituts für industrielle Technologien und Wissenschaften der Fachhochschule Westschweiz (HES-SO Genf/HEPIA) wurde ein Prototyp eines einfach zu handhabenden und transportablen Messinstruments und des dazugehörigen Messverfahrens entwickelt und im Oktober 2020 zum Patent angemeldet (Impact Paragraph). Mit dem Ziel, das angepasste Messverfahren zu evaluieren, um eine spätere Anwendung im klinischen Alltag zu ermöglichen, wurde das Studienprotokoll einer Reliabilitäts- und Validitätsstudie mit dem angepassten Messinstrument und -verfahren entwickelt und wird in Kapitel 6 näher erläutert.

Teil 2: Die diagnostische Genauigkeit der derzeit verwendeten Sturzrisikotests (einschliesslich Kraftmessungen) hinsichtlich Erkennung und/oder Vorhersage des Sturzrisikos

In **Kapitel 7** wurden in einer systematischen Übersichtsarbeit die klinimetrischen Eigenschaften anderer Tests, welche die International Classification of Functioning, Disability and Health (ICF) Komponenten „Körperfunktionen und -strukturen“ und/oder "Aktivitäten und Partizipation" bewerten, in Bezug auf das Sturzrisiko älterer Menschen untersucht. Im Bewusstsein der multifaktoriellen Ursache von Stürzen und unter der Annahme, dass für eine gute Sturzrisikoerkennung die Kraft der Hüftabduktoren mit anderen leistungsbasierten Sturzrisikobewertungen kombiniert werden muss, wollten wir einen Überblick über die Reliabilität (Inter- und Intra-Rater) und die diagnostische Genauigkeit, insbesondere die Post-Test-Wahrscheinlichkeit, von häufig verwendeten Leistungstests in Abhängigkeit von ihrer Umgebung (selbständig zuhause lebend, Pflegeheim oder Krankenhaus) geben. Ziel war es, dass diese Übersicht das Gesundheitspersonal bei der Entscheidungsfindung zusammen mit der sturzgefährdeten Person unterstützen kann, um bei der Auswahl eines geeigneten Sturzrisikotests entsprechend der spezifischen Umgebung der sturzgefährdeten Person zu helfen. Insgesamt wurden 42 Artikel für das systematische Review ausgewählt, von denen 21 Artikel Angaben zur Reliabilität und 25 zur Validität enthielten. Vier der eingeschlossenen Artikel enthielten Informationen zur Reliabilität und zur Validität des eingeschlossenen

Tests. Um die klinische Anwendbarkeit zu erleichtern und weil die diagnostische Genauigkeit der Tests stark von der untersuchten Population abhängt, haben wir unsere Ergebnisse nach den verschiedenen Umgebungen (ambulant, im Pflegeheim, im Krankenhaus und in einer kombinierten Umgebung aus stationären und ambulanten Personen) zusammengefasst. Im Allgemeinen zeigte das systematische Review, dass die Unterscheidungsfähigkeit der einbezogenen Tests zur Erkennung von sturzgefährdeten Personen für den klinischen Einsatz nicht sehr gut ist. Die meisten der einbezogenen Tests zeigen nur relativ niedrige Sensitivitätswerte und geringe Veränderungen zwischen den Wahrscheinlichkeiten vor und nach dem Test ein Sturzrisiko zu erkennen. Unsere Ergebnisse sowie das Wissen um die multifaktorielle Ursache von Stürzen unterstreichen, dass die derzeit verwendeten Sturzrisikobewertungen bei einmaliger Anwendung nicht gut genug sind, um zwischen sturzgefährdeten und nicht-sturzgefährdeten älteren Personen zu unterscheiden. Um eine geeignete Sturzrisikobewertung auswählen zu können, sollten Fachkräfte im Gesundheitswesen die klinimetrischen Eigenschaften eines Tests kennen und verstehen. In unserer systematischen Übersichtsarbeit erläutern wir für Kliniker*innen wichtige Aspekte eines Messinstruments, ihre Einflussfaktoren und ihre Interpretation. Eine klinische Anwendung der Ergebnisse in der täglichen Praxis ist deshalb leicht möglich und unterstützt den Entscheidungsprozess zum Verlauf der weiteren Therapie zwischen Patient*in und Gesundheitspersonal.

Kapitel 8 und Kapitel 9 schliessen diese Arbeit mit einer allgemeinen Diskussion und dem Abschnitt über die Auswirkungen der Forschungsergebnisse ab. In **Kapitel 8** werden unsere wichtigsten Ergebnisse erörtert, Überlegungen zu den Stärken, Grenzen und methodischen Aspekten unserer Studien gemacht, und Implikationen für die klinische Praxis und künftige Forschung aufgezeigt. Die Messung der Kraft der Hüftabduktoren trägt zum Verständnis des Sturzrisikos älterer Menschen bei. Ihre Evaluierung ist bei älteren Menschen durchführbar, weist gute klinische Eigenschaften auf und bietet einen zusätzlichen Wert zu den derzeit verwendeten Tests des Sturzrisikos und kann somit auch zu einer Verbesserung von Sturzpräventions- und Rehabilitationsprogrammen beitragen. Die Kraft der Hüftabduktoren bewertet ergänzende Aspekte von Stürzen, wird aber bisher noch nicht systematisch gemessen. Darüber hinaus zeigen die derzeit verwendeten Sturzrisikotests nur eine mässige Fähigkeit zur Unterscheidung zwischen sturzgefährdeten und nicht-sturzgefährdeten Personen. Zudem hat die Umgebung (ambulant, im Pflegeheim, im Spital, etc.), in welcher die Tests durchgeführt werden, einen grossen Einfluss darauf, welche Tests für den klinischen Einsatz empfohlen werden können. Die klinische Anwendbarkeit der Ergebnisse wird in Kapitel 8 hervorgehoben. Es wird darauf hingewiesen, dass die Forschenden ihre Arbeit für die direkte Anwendung im klinischen Alltag verständlich machen müssen, damit

die Lücke zwischen Forschung und Praxis geschlossen und Leitlinien für die bestmögliche Behandlung der Patient*innen entwickelt werden können.

In **Kapitel 9** wird die wissenschaftliche und gesellschaftliche Bedeutung meiner Forschungsprojekte und vor allem die Bedeutung der Ergebnisse für die Patient*innen beleuchtet. Die vorliegenden Forschungsarbeiten zielen darauf ab, die klinische Arbeit und damit die Ergebnisse für die Patient*innen zu verbessern, indem relevante Parameter für die künftige Entwicklung einer gezielten Therapie des Sturzrisikos ermittelt werden. Die aus den verschiedenen Projekten dieser Doktorarbeit gewonnenen Erkenntnisse sind für Forschende, Gesundheitspersonal, Personen im Alter von ≥ 65 Jahren, und Angehörige von sturzgefährdeten Personen von Nutzen. Sie sind daher von grosser wissenschaftlicher, gesellschaftlicher und klinischer Relevanz. Zu diagnostischen Zwecken empfehlen wir die Erfassung der Kraft der Hüftabduktoren und ermutigen Klinker*innen, die entsprechende Messung im täglichen klinischen Gebrauch zu integrieren.

RÉSUMÉ

Cette thèse avait pour but de mieux comprendre le rôle de la force des muscles abducteurs de la hanche dans le risque de chute chez les personnes âgées et d'évaluer si cette force permet de distinguer les personnes âgées qui chutent de celles qui ne chutent pas. Nous nous sommes intéressés à la possibilité d'utiliser la force des muscles abducteurs de la hanche pour évaluer le risque de chute, soit seule ou au sein d'une batterie de tests.

La **préface** décrit un scénario que nous avons tous vécu. Il nous fait prendre conscience de l'importance des chutes et des conséquences délétères qui leur sont associées. Cela permet une entrée émotionnelle dans la recherche pour bien souligner que ce projet n'était pas simplement "un de ces projets de plus sur l'évaluation des risques de chute...". La préface nous fait réaliser à quel point ce sujet peut nous toucher personnellement et à quel point nous lui sommes émotionnellement liés.

- Qu'aimeriez-vous savoir pour aider vos proches à prévenir une chute?
- Quelles sont les questions auxquelles vous aimeriez obtenir une réponse en ce qui concerne le risque de chute?
 - o Quels sont les facteurs qui influencent les chutes?
 - o Quels tests permettent d'évaluer les facteurs qui influencent les chutes?
 - o Que dit la science au sujet du risque de chute et de son évaluation?

Le **chapitre 1** présente l'importance du risque de chute. Il souligne les changements démographiques qui conduisent à une augmentation considérable du nombre de personnes âgées et, subséquemment, à une augmentation de l'incidence des chutes. De nombreux facteurs influencent le concept de "risque de chute". Cependant, cette thèse se concentre sur les facteurs physiologiques, les paramètres qui sont modifiables et importants pour guider le contenu d'une intervention de physiothérapie, tels que la diminution du contrôle postural, de la force et de la mobilité. Elle aborde aussi les aspects que nous devons impérativement comprendre tels que les facteurs qui influencent les chutes, les types de tests disponibles pour évaluer ces facteurs, l'influence de la force musculaire des membres inférieurs - et plus particulièrement l'intérêt pour la force des muscles abducteurs de la hanche en relation avec le risque de chute. Ce chapitre donne des informations sur les paramètres importants qui guident le choix d'un test d'évaluation du risque de chute ainsi que sur la taxonomie, les définitions et les aspects méthodologiques des études clinimétriques. En fait, la faiblesse fonctionnelle des membres inférieurs chez les personnes âgées est un indicateur fort. Elle pourrait favoriser la transition malencontreuse d'une personne qui ne chute pas à une personne qui chute; son

évaluation et son traitement chez les personnes à risque de chute sont recommandés. La littérature récente souligne l'importance de la force des muscles abducteurs de la hanche pour maintenir le contrôle postural médiolatéral, un facteur qui est souvent altéré chez les personnes âgées, en particulier chez celles qui chutent. La force des muscles abducteurs de la hanche dans le plan frontal et les moments de force spinaux sont les paramètres majeurs pour contrecarrer la déstabilisation dans la direction médiolatérale pendant l'appui sur une seule jambe. Le déclin de la force de ces muscles est plus prononcé que celui des muscles qui agissent dans le plan sagittal. Les faits mentionnés nous ont conduit à formuler l'hypothèse que la force des muscles abducteurs de la hanche pourrait être un facteur contributif prometteur pour évaluer le risque de chute. Comme cette approche semblait être cliniquement pertinente et nouvelle, nous avons décidé d'étudier le groupe des muscles abducteurs de la hanche plus en détail. Nous avons prévu d'établir des valeurs-seuils et une précision diagnostique afin de faciliter le processus de décision clinique concernant le risque de chute des personnes âgées. La validation de cette approche fournit aux cliniciens une évaluation supplémentaire à la batterie des tests du risque de chute déjà existante, mais elle pourrait également fournir des informations cruciales sur un paramètre directement ciblé pour la prévention des chutes.

Dans cette thèse, nous avons abordé deux questions principales; elles sont discutées dans la Partie 1 et la Partie 2.

Partie 1: Rôle de la force des muscles abducteurs de la hanche dans le domaine de l'évaluation du risque de chute et développement d'un nouvel instrument de mesure ainsi que d'une nouvelle procédure pour la clinique quotidienne.

La première partie de cette thèse (chapitres 2 à 6) s'est concentrée sur le rôle de la force des muscles abducteurs de la hanche dans l'évaluation du risque de chute des personnes âgées de 65 ans et plus. Il était prévu d'évaluer la faisabilité, la fiabilité et la validité de notre méthode de mesure de la force dans le plan frontal de la hanche. Dans un projet ultérieur, nous avons développé un test de force des muscles abducteurs de la hanche afin d'obtenir une nouvelle procédure de test fonctionnelle et facile à utiliser.

Dans les **chapitres 2 et 3**, la faisabilité et la fiabilité intra-évaluateur de la mesure de la force des muscles abducteurs de la hanche ont été évaluées dans le plan frontal de la hanche en position couchée sur le côté et en position debout. La fiabilité relative et absolue est une condition préalable à une bonne validité. Dans le **chapitre 2**, nous avons utilisé une méthode standardisée pour mesurer

la force des muscles abducteurs de la hanche et des muscles adducteurs de la hanche : le sujet était en position couchée sur le côté et poussait sa jambe aussi fort et rapidement que possible contre un dynamomètre fixé à un cadre métallique fabriqué sur mesure et fixé à la table de traitement. Soixante-seize personnes, âgées de 65 ans et plus, ont été recrutées dans un hôpital gériatrique qui fait partie des Hôpitaux universitaires de Genève et dans un cabinet de consultation externe en Suisse. Trente-huit personnes ont subi le test d'abduction et les 38 autres le test d'adduction. Les participants ont été testés deux fois dans un délai d'une à trois heures par le même évaluateur. Le coefficient de corrélation intraclasse, l'erreur standard de mesure et la plus petite différence détectable ont été déterminés. Il en ressort que la force isométrique maximale volontaire et que le taux de génération de force des muscles abducteurs de la hanche et des muscles adducteurs de la hanche en position couchée sur le côté est réalisable et présente des valeurs élevées de fiabilité chez les personnes de 65 ans et plus. Cependant, le temps d'évaluation - environ 10 minutes - en position couchée sur le côté, est trop long pour utiliser ce test en clinique quotidienne.

Cela nous a conduits à évaluer la faisabilité et la fiabilité de la mesure de la force des muscles abducteurs de la hanche en position debout, décrite au **chapitre 3**. Comme pour les mesures décrites dans le chapitre 2, des personnes âgées de 65 ans et plus ont été recrutées dans le service de gériatrie d'un hôpital universitaire et dans une clinique externe. Trente-deux personnes, dont 16 chuteurs (≥ 1 chute au cours des 12 derniers mois) et 16 non-chuteurs ont été inclus. La force isométrique maximale volontaire et le taux de génération de force des muscles abducteurs de la hanche de la jambe droite ont été mesurés en position debout à l'aide d'un dynamomètre manuel. Deux sessions de mesure ont été effectuées par le même évaluateur à une ou trois heures d'intervalle. La fiabilité a été déterminée en utilisant le coefficient de corrélation intraclasse, l'erreur standard de mesure et une analyse de Bland et Altman. Les résultats ont montré que la mesure de la force des muscles abducteurs de la hanche en position debout est faisable; elle atteint des valeurs d'accord de coefficient intraclasse aussi élevées que celles des mesures en position couchée sur le côté. L'avantage du test en position debout est qu'il prend environ la moitié moins de temps que le test en position couchée sur le côté. Il semble donc beaucoup plus réalisable en pratique clinique. Bien que les deux positions de test aient montré une bonne fiabilité intra-évaluateur, des erreurs standards de mesure et des plus petites différences détectables élevées ont été enregistrées dans les deux cas, avec des valeurs de plus petite différence détectable légèrement inférieures en position debout qu'en position couchée sur le côté.

Afin de répondre à la question de l'importance de la force des muscles abducteurs de la hanche par rapport au risque de chute, nous avons évalué au

chapitre 4 lequel des six groupes musculaires de la hanche permet le mieux de distinguer les personnes âgées qui chutent de celles qui ne chutent pas et nous avons comparé leur capacité de discrimination avec celle de la force de préhension de la main. Chez 60 personnes âgées de 65 ans et plus (38 femmes, 22 hommes), la force isométrique maximale volontaire et le taux de génération de force des muscles abducteurs, adducteurs, rotateurs internes, rotateurs externes, extenseurs et fléchisseurs de la hanche ont été mesurés à l'aide d'un dynamomètre fixé à un cadre fabriqué sur mesure; la force de préhension de la main a aussi été mesurée, à l'aide d'un Vigorimètre Martin. Par rapport à la norme de référence prédéfinie, l'historique des chutes, il s'est avéré que la force des muscles abducteurs de la hanche était le seul groupe musculaire de la hanche qui présentait une capacité discriminante suffisamment bonne pour une utilisation clinique (aire sous la courbe (AUC) $> 0,7$). Les résultats du chapitre 4 indiquent que la force des muscles abducteurs de la hanche est une mesure utile pour distinguer les adultes âgés de 65 ans et plus qui chutent de ceux qui ne chutent pas en ce qui concerne le critère externe choisi.

Cela nous a conduits à poursuivre l'évaluation des propriétés psychométriques spécifiques de la force des muscles abducteurs de hanche. Le **chapitre 5** présente une évaluation étendue et une étape suivante des résultats du chapitre 4. La précision diagnostique de la force des muscles abducteurs de la hanche a été examinée afin de décider si ce paramètre doit être intégré dans une batterie de tests pour évaluer les risques de chute des personnes âgées. Pour cela, nous avons calculé la sensibilité et la spécificité nettes afin d'avoir un aperçu de la combinaison possible de la force des muscles abducteurs de hanche avec les évaluations du risque de chute que nous avons choisies, à savoir la *Short Physical Performance Battery* (SPPB) et le test *Timed up and Go* (TUG). Les données démographiques des 60 personnes incluses (par exemple, l'âge, le sexe et l'IMC) ont été extraites de la base de données clinique; les antécédents de chute, standard de référence, ont été recueillis par un interrogatoire verbal, puis des physiothérapeutes formés (en aveugle par rapport aux antécédents de chute des participants) ont effectué la *Falls Efficacy Scale International*, la *Mini Mental State Examination*, le SPPB, le TUG et mesuré la force des muscles abducteurs de la hanche des participants (force isométrique maximale volontaire et taux de génération de force). Les résultats du chapitre 5 indiquent que la force des muscles abducteurs de la hanche (force isométrique maximale volontaire et taux de génération de force) présente une bonne précision diagnostique pour différencier les personnes âgées qui chutent de celles qui ne chutent pas. Lorsque les mesures de la force des muscles abducteurs de la hanche (force isométrique maximale volontaire et taux de génération de force) sont combinées avec les évaluations du risque de chute actuellement utilisées, la sensibilité nette élevée montre une valeur contributive prometteuse dans une batterie

de tests. La force des muscles abducteurs de la hanche est un paramètre facilement mesurable et devrait, sur une base clinique quotidienne, être considérée comme un facteur d'influence important du risque de chute.

Cependant, au cours des différents projets des chapitres 2 à 5, nous avons été confrontés à de nombreux défis en ce qui concerne la position d'évaluation et l'instrument de mesure utilisés pour mesurer la force des muscles abducteurs de la hanche ainsi qu'en ce qui concerne son applicabilité fonctionnelle et surtout clinique. Cela nous a conduits à réfléchir et à trouver un moyen d'évaluer la force des muscles abducteurs de la hanche dans la vie clinique quotidienne. Par conséquent, le projet supplémentaire de développement d'une procédure de mesure en position debout a émergé afin de permettre l'intégration de nos résultats précédents dans la pratique clinique. Ce projet est abordé **au chapitre 6**. Avec l'aide d'ingénieurs de l'*Institut des technologies et des sciences industrielles* de la *Haute école spécialisée de Suisse occidentale* (HES-SO Genève/HEPIA), un prototype d'un instrument de mesure facile à manipuler et à transporter ainsi qu'une procédure de mesure associée ont été développés; un brevet a été déposé en octobre 2020 (paragraphe de l'impact). Un protocole d'étude de la fiabilité et de la validité de l'instrument et de la procédure de mesure été développé dans le but d'évaluer la procédure de mesure adaptée pour une utilisation ultérieure dans la vie clinique quotidienne.

Partie 2: Exactitude diagnostique des évaluations du risque de chute actuellement utilisées - y compris les mesures de force - afin d'avoir un aperçu de leur capacité à détecter et/ou à prédire le risque de chute.

Le **chapitre 7** présente la revue systématique effectuée pour évaluer les propriétés clinimétriques d'autres tests qui évaluent les fonctions et structures corporelles et/ou les composantes «activités et participation» de la Classification internationale du fonctionnement, du handicap et de la santé (CIF) en relation avec le risque de chute chez les personnes âgées. Conscients de l'étiologie multifactorielle des chutes et supposant que la force des muscles abducteurs de la hanche devra être combinée avec d'autres évaluations du risque de chute basées sur la performance pour bien détecter le risque de chute, nous avons cherché à fournir un aperçu de la fiabilité (inter- et intra-évaluateur) et de la précision diagnostique, en particulier la probabilité post-test des tests de performance couramment utilisés pour évaluer le risque de chute des personnes âgées de 65 ans et plus, en fonction de leur contexte (logement communautaire, institution de soins ou hôpital). L'objectif était que cette vue d'ensemble puisse aider les soignants à choisir une évaluation du risque de chute appropriée et qui corresponde au cadre spécifique de la personne à risque de chute pour ensuite prendre des décisions avec elle.

Quarante-deux articles ont été retenus pour la revue systématique; 21 contenaient des informations sur la fiabilité, 25 sur la validité; 4 contenaient des informations sur la fiabilité et la validité. Afin de faciliter l'applicabilité clinique et parce que la précision diagnostique des tests est fortement influencée par la population étudiée, nous avons résumé les résultats en fonction du contexte dans lequel ils ont été obtenus (personnes vivant à leur domicile, en institution de soins, hôpital et contexte combiné de personnes hospitalisées et de personnes vivant à leur domicile). La revue systématique a montré que la capacité de discrimination des tests pour détecter les personnes de 65 ans et plus à risque de chute n'est pas très bonne pour une utilisation clinique. La plupart des tests ne présentent que des valeurs de sensibilité relativement faibles et de faibles changements entre les probabilités avant et après le test. Ces résultats, ainsi que la connaissance de l'étiologie multifactorielle des chutes, soulignent que les évaluations du risque de chute actuellement utilisées ne sont pas suffisamment bonnes pour distinguer les personnes âgées de 65 ans et plus qui chutent de celles qui ne chutent pas. Pour être en mesure de choisir une évaluation du risque de chute appropriée, les professionnels de santé doivent connaître et comprendre les propriétés clinimétriques d'un test. Dans notre revue systématique, nous expliquons les aspects importants d'un instrument de mesure, leurs facteurs d'influence et leur interprétation pour les cliniciens. Il est facile d'appliquer ces résultats dans la pratique quotidienne; ils soutiennent le processus de décision entre le patient et le soignant.

Les **chapitres 8 et 9** complètent cette thèse par une discussion générale et un paragraphe d'impact. Le **chapitre 8** discute les principaux résultats, présente des réflexions sur les forces, les limites et les aspects méthodologiques des études que nous avons effectuées. Il propose aussi des implications pour la pratique clinique et les recherches futures. La mesure de la force des muscles abducteurs de la hanche aide à comprendre les risques de chute des personnes âgées. Elle est réalisable chez les personnes âgées, présente de bonnes propriétés clinimétriques et apporte une valeur ajoutée aux évaluations du risque de chute actuellement utilisées et, par conséquent, aux programmes de prévention du risque de chute et de réadaptation. Elle éclaire des aspects complémentaires des chutes, mais n'est pas encore mesurée systématiquement. Les évaluations du risque de chute basées sur la performance actuellement utilisées ne montrent qu'une capacité modérée distinguer les personnes qui chutent de celles qui ne chutent pas. Le contexte dans lequel les tests sont effectués influence fortement le type de tests qui peuvent être recommandés en clinique quotidienne. L'applicabilité clinique des résultats est soulignée.

Il est important que les chercheurs veillent à rendre leurs travaux compréhensibles et directement applicables dans la clinique quotidienne. Il est aussi

nécessaire de combler le fossé entre la recherche et la pratique afin de développer des recommandations de bonne pratique pour les patients.

Enfin, le **chapitre 9** propose une réflexion sur la signification scientifique et sociétale de ces projets de recherche et, surtout, sur la signification des résultats au niveau des patients. La présente recherche visait à améliorer le travail clinique et les résultats des patients en identifiant les paramètres pertinents pour le développement futur d'un traitement ciblé pour le risque de chute.

Les connaissances acquises dans le cadre des différents projets qui composent cette thèse de doctorat sont bénéfiques aux chercheurs, aux professionnels de santé, aux personnes âgées ≥ 65 ans, aux soignants et aux proches des personnes à risque de chute. Elles ont donc un fort impact scientifique, sociétal et clinique. Nous recommandons d'évaluer la force des muscles abducteurs de la hanche à des fins diagnostiques et encourageons son évaluation dans la pratique clinique quotidienne.

ACKNOWLEDGMENTS

The journey on the way to finish a doctoral thesis is sometimes described as a solitary endeavor; however, I think that the long list that follows definitely proves the opposite.

I would like to express my gratitude and thanks to all the persons that accompanied me, one way or the other, on my PhD journey and during the development of my way of seeing things and thinking. It has been a long and at times not that easy challenge, which is why I am even happier that you all were there for me during that time by providing feedback, support, help, inspiration, and most importantly encouragements as well as distractions.

Firstly, I would like to express my sincere gratitude to my promotion team, without you I would have never been able to carry out my research projects, been able to publish my articles and write my thesis.

Caroline Bastiaenen, thank you for your guidance throughout my whole PhD journey, which helped me during the research projects and writing of the articles and my thesis. Your wide knowledge of research methodologies, diagnostic studies and research in general were always a big help. My two stays in Maastricht were additionally a great opportunity to directly work with you and benefit even more from your expertise. Thank you so much for all the things I could learn from you and the valuable baggage you gave me for my way in academia.

Rob de Bie, we did not meet in person very often but every time we met, you were able to rapidly understand the current challenges of my projects and give very helpful advice. With only a few sentences, you were able to reformulate the problem and always asked the right questions. You made it look easy to find a way out of some dead ends and challenging labyrinths my PhD projects faced during the last seven years. Thank you very much for your support.

Lara de Preux-Allet, as my daily supervisor you were the most present person during my whole PhD trajectory. Thank you very much for your continuous support, your patience, motivation, and knowledge for all PhD related projects as well as all the other research projects for which you opened the doors for me. During all the time you accompanied me, from my first steps in research during my master thesis to now finishing my doctoral thesis, we experienced a lot together, be it spinning classes, (gala-) dinners, conferences, and many discussions about research topics and personal things. It was an eventful timespan for both of us, as for example your three children were born during that time of which one kind of woke you up during the night at my apartment to let you know it was time to go to the hospital. Lara I thank you very much for all you have done for me during the last ten years, I am looking forward to many constructive years to come.

Besides my supervisor team, I would like to thank the assessment committee: **Prof. dr. JMGA Schols (chair), Prof. dr. ED de Bruin, Dr. JSM Hobbelen, Prof. dr. AF Lenssen and Dr. C. McCrum** to have thoroughly read my thesis and your insightful comments.

Many thanks go to all persons that were implicated in my research projects.

Roger Hilfiker, thank you for the numerous hours you spent on the statistics of my research projects and the even greater number of hours you spent explaining them to me. Maybe you partially were at the origin of when the cornerstone of my career in research was laid, as a long time ago you were the supervisor of my Bachelor thesis. I am grateful for all the work you have done for my thesis and me. Thank you very much.

Philippe Terrier, you were there at the beginning of my research projects, the moment I very excitedly gathered my first data and had to perform the first analyses. It is always easy to forget at the end of a project, that the beginning of the data treatment went through very challenging phases at times. Thank you for your patience and help.

Prof. dr. Gabriel Gold and Prof. dr. Serge Ferrari, thank you to have allowed me to perform our research projects with the patient recruitment in your departments at the University Hospitals of Geneva. Your valuable feedback and inputs during the development phase of the methodology of the research projects and the article writing were very much appreciated.

Dominique Monnin, thank you that you are accompanying me as a mentor. You were a very important, calm cornerstone at the challenging phase of getting the PhD research projects started during the maternity leave of Lara. Your encouragement helped me overcome some difficult phases. Your foresight and experience in managing projects were always very helpful. I am honored to have been able to learn a lot from you. A very big thank you for all the French translations and corrections you did for me, as well as your patience when you once again had to read a “bref” (short) three-page long e-mail from me. I am happy that you were always there for me during the PhD journey and will be there for me in many projects to come. Thank you very much!

Ilona Punt, you were in Geneva during my master thesis and the first years of my PhD. Thank you very much for all your advice, feedback for my articles and grant applications as well as your help with some of the measurements. It was very nice to meet you again during some dinners and lunches in Maastricht during my two stays.

A very big thank you to **all the physiotherapists of the University Hospitals of Geneva** that helped during the patient recruitment and various measurements. Thank you for all the great work you do every day. Without you, working on my

different research projects and finishing my thesis would have been even more challenging and a lot more time-consuming.

For sure, a very big thank you goes to **all the participants** of my research projects. Without your willingness to perform the sometimes very time-consuming measurements, it would have been impossible to complete this dissertation. THANK YOU

Thank you also to the many **last-minute helpers** for grant and article submissions and all other things that suddenly hurried; **Stéphanie Martin-Christie, Nathalie Bourquin Berger, Sabrina Simi, Clara James, Yohan Van de Looij, Derek Christie**, and all other persons I might forget.

Thank you also to all the **teachers and research colleagues** at the University of Applied Sciences and Arts of Western Switzerland, Geneva, that helped me in one way or the other. It was a challenging time for me, your support was a big help. Thank you.

A very big THANK YOU goes to all my friends and family who supported me during all this time.

Judith Graser, you are my favorite Swiss-paranymph. We started our PhD's almost at the same time and since then had many exchanges by mail, a few phone calls, and direct meetings where we usually found a little time to complain, but also exchanged important information, encouraged each other but most importantly always found the time to laugh together and cheer for each other. I am happy you were there for me when I needed to hear that I am still on the right path. I am looking forward to your encouraging words before and your emotional support during my defense. Thank you!

Miriam Weijers, we met during my stay in Maastricht in 2019 and now you are my favorite Dutch-paranymph now. Your calm and very positive way to see the world and enjoy all the small things in life is very inspiring to me. I am very happy we could spend some time together while I was in Maastricht and that the contact continued after I went back to Switzerland. You were at the right time at the right place, I am very grateful for that. Thank you that you will also be there for me during the defense and provide some positive emotional support.

Anna Henzi, thank you very much for always have had an open ear for all my calls for help. Our little helpline eased many (mostly unnecessary) worries. It was nice to have somebody to talk to that really understands and already went through some PhD struggles as well. Thank you very much for all the interesting discussions and adventures we had together and of which we will hopefully have a lot more in the future. Thank you as well for the emergency survival packages that unexpectedly arrived at my door from time to time. Thank you very much.

Pascale Brand, we met at the physiotherapy school a long time ago and since then kind of did all our physiotherapy education together. It was always a delight to spend time together with you, be it during dinners, brunches, or help you to move in or out of apartments but mostly during nice outdoor adventures. Your way of seeing the world often motivated me to reflect. Thank you very much as well, that I could spend some resourcing time at your peaceful, splendid home. A thank you also goes to **Roger Rathgeb** who gave me feedback to the one incorrect comma in the German summary..., and who was a great host with all the nice cooking, discussions, and made many nice fires to heat the beautiful sauna. I am looking forward to spending a lot more adventures with you two as well as to encourage and cheer for you during your PhD time Pascale. A big thank you goes to both of you.

Stéphanie Saxer, thank you for always have listened to my sorrows and given me helpful advice during the many times you called me. Thank you as well for all the nice summaries you made during the master and shared with me. I am sure they also helped to lay the foundation stones allowing me to now finish my dissertation. Thank you also for all the fun stuff we did together like going to Europa Park or dancing at your wedding.

Simone Elsig my co-Maastricht-PhD-colleague, thank you for the nice discussions we had during the different PhD meetings. You helped me a lot with your calm nature and the always helpful advice. Thank you that you were there for me when I needed some encouragement when once again the projects progressed more slowly than anticipated or suddenly changed directions.

Tamara Hofer, Nadja Studer, Nicole Stump, my friends from a very long time ago. It is nice to know you since our time at the high school and still exchange some messages and occasionally visit each other. Although we do not meet much it is always as if the time stood still since the last time we saw each other and the discussions take off as if there was no separation. Many years ago we attended your PhD defence and I admired you for it, Tamara, although at the time I did not know what it really meant to finish a dissertation. Now I know... I really would like to thank you all for the nice moments we share together now and then.

Marja de Graaf- van den Heuvel, Koen de Graaf and Kees Brinkhof, you are some more important Dutch people in my life. You were there since the beginning of my journey as a physiotherapist. My very first job after physiotherapy school was in your practice. Later you gave me the opportunity to do my Master's degree while working in your practice, therefore, allowing me to develop and open "my wings to fly" and follow my own path in the world of physiotherapy. Thank you very much.

To all **my friends** that shared their valuable time with me during this long period and I might forget to mention separately, THANK YOU. It is the time with you that brought the much-needed distractions and happy moments away from my research

world. Without you, I would not have been able to finish my dissertation. Thank you very much.

A thank you also goes to the **colleagues of my swimming teams**, although you do not know it, you were an important part of my life during my PhD. Sometimes it does not need more than an hour of splashing around with water and making silly jokes to make a difference in a day. Thank you to have been my companions through these times.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help, and support.

Mami and Papi, my beloved parents, thank you for all your encouragement and support since my first day, and a big thank you for having been there for me during this challenging time of my PhD journey. You taught me to be curious and explore things and gave me the much-needed fundamentals to dream big and reach for the stars but also being able to work hard and achieve my goals. THANK YOU that you made it possible for me to make the experiences that have made me who I am. This journey would not have been possible without you. This milestone, therefore, is dedicated to you.

Mami und Papi, meine lieben Eltern, ich danke euch für all eure Ermutigungen und eure Unterstützung seit meinem ersten Tag und ein großes Dankeschön dafür, dass ihr in dieser herausfordernden Zeit meiner "PhD-Reise" für mich da wart. Ihr habt mir beigebracht, neugierig zu sein und Dinge zu erforschen, und ihr habt mir das dringend benötigte Fundament gegeben, um große Träume zu haben und nach den Sternen zu greifen, aber auch, um hart zu arbeiten und meine Ziele zu erreichen. DANKE, dass ihr mir ermöglicht habt, die Erfahrungen zu machen, die mich zu dem gemacht haben, was ich heute bin. Diese Reise wäre ohne euch nicht möglich gewesen. Dieser Meilenstein ist deshalb euch gewidmet.

Fabienne Gafner, my "little" sister, I am grateful that you are always there for me when I need it; I know that I always can count on you. All the nice little gifts and the survival kit packages that occasionally arrived at my home were always a great joy. Although the packages were huge, the nerve-calming sweets did not usually last long... Thank you for your help with cross-reading some of my texts and presentations. The biggest thank you, however, goes to you for encouraging me in all of my pursuits and inspiring me to follow my dreams. A thank you also goes to **Matthias Felix** for the nice discussions and the right questions to help me figure out some answers and facilitating taking some decisions. The work on the steep slope of your garden also was a nice distraction and helped to empty some space in the overloaded head.

I thank all of you for sharing part of my adventurous journey with me.

THANK YOU

CURRICULUM VITAE

Simone Chantal Gafner was born on April 14th, 1983 in Bern, Switzerland. She obtained in 2007 her Bachelor of Science in Physiotherapy at the HES-SO University of Applied Sciences and Arts of Western Switzerland, Leukerbad, Valais-Wallis.

After 4 years as a full-time physiotherapist in different clinical practices in the French and German speaking part of Switzerland and various continues educations mostly in the musculoskeletal field, such as manual therapy and sports physiotherapy, she started a part time Master of Science in physiotherapy at the Zurich University of Applied Sciences, School of Health Professions, Institute of Physiotherapy, in Winterthur, Switzerland. Next to the master Simone continued her practical work as a physiotherapist in a private practice in Schuepfen (Bern). In 2014 Simone completed her master's degree with a thesis on the influence of hip abductor fatigue on the knee frontal plane angles during single-leg landings and side-step cutting maneuvers.

Thereafter Simone had the chance of pursuing her research activity. She started her PhD trajectory as an external doctoral candidate at the Department of Epidemiology, Research Line Functioning, Participation and Rehabilitation, CAPHRI at the Maastricht University in the Netherlands in collaboration with the Geneva School of Health Sciences, HES-SO University of Applied Sciences and Arts Western Switzerland, in Geneva. Hip abductor strength stayed the main topic of her research, although she now focused on persons aged 65 years and over and their risk of falls, under the local, daily supervision of Prof. Dr. Lara Allet and under supervision of Dr. Caroline Bastiaenen and Prof. Dr. Rob de Bie.

Simone obtained two main grants that allowed her to conduct her research projects. The first grant, a « PhD grant », was awarded by the foundation "Physiotherapie Wissenschaften". A second grant was awarded by the University of Applied Sciences and Arts Western Switzerland (Commission scientifique du domaine Santé; HES-SO fond de recherche et impulsion) for her systematic review "Diagnostic accuracy of commonly-used performance tests relative to fall history in older persons: a systematic review". Additionally Simone was awarded two prizes; first the Elisabeth Buerge Price for the best first scientific article and second the research award of the rehabilitation clinic "Reha Rheinfelden" for her article "Hip muscle and hand-grip strength to differentiate between older fallers and non-fallers: a cross-sectional validity study".

LIST OF PUBLICATIONS

Gafner SC, Allet L, Hilfiker R, Bastiaenen CHG. Reliability and Diagnostic Accuracy of Commonly Used Performance Tests Relative to Fall History in Older Persons: A Systematic Review. *Clinical interventions in aging*. 2021;16:1591-1616.

Reinmann A, **Gafner SC**, Hilfiker R, Bruyneel AV, Pataky Z, Allet L. Bariatric Surgery: Consequences on Functional Capacities in Patients With Obesity. *Front Endocrinol (Lausanne)*. 2021;12:646283.

Gafner SC, Bastiaenen CHG, Biver E, Ferrari S, Allet L. Reliability and validity of an adapted hip abductor strength measure as a potential new fall risk assessment for older persons: a study protocol. *BMC geriatrics*. 2021;21(1):110.

Gafner SC, Bastiaenen CHG, Ferrari S, et al. The Role of Hip Abductor Strength in Identifying Older Persons at Risk of Falls: A Diagnostic Accuracy Study. *Clinical interventions in aging*. 2020;15:645-654.

Gafner SC, Hoevel V, Punt IM, Schmid S, Armand S, Allet L. Hip-abductor fatigue influences sagittal plane ankle kinematics and shank muscle activity during a single-leg forward jump. *J Electromyogr Kinesiol*. 2018 Sep 12;43:75-81. doi: 10.1016/j.jelekin.2018.09.004. ¹ PubMed PMID: 30243233.

Gafner SC, Bastiaenen CHG, Ferrari S, Gold G, Terrier P, Hilfiker R, Allet L. Hip muscle and hand-grip strength to differentiate between older fallers and non-fallers: a cross-sectional validity study. *Clinical interventions in aging*. 2018;13:1-8.

Bruyneel AV, **Gafner SC**, Ferrari S, Gold G, Monnin D, Terrier P, Bastiaenen CHG, Allet L. Intra-rater reliability of hip abductor isometric strength testing in a standing position in older fallers and non-fallers. *European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity*. 2018;15:9.

Gafner S, Bastiaenen CHG, Terrier P, Punt I, Ferrari S, Gold G, de Bie R, Allet L. Evaluation of hip abductor and adductor strength in the elderly: a reliability study. *European review of aging and physical activity: official journal of the European Group for Research into Elderly and Physical Activity*. 2017;14:5.